

Methods 4 - 5

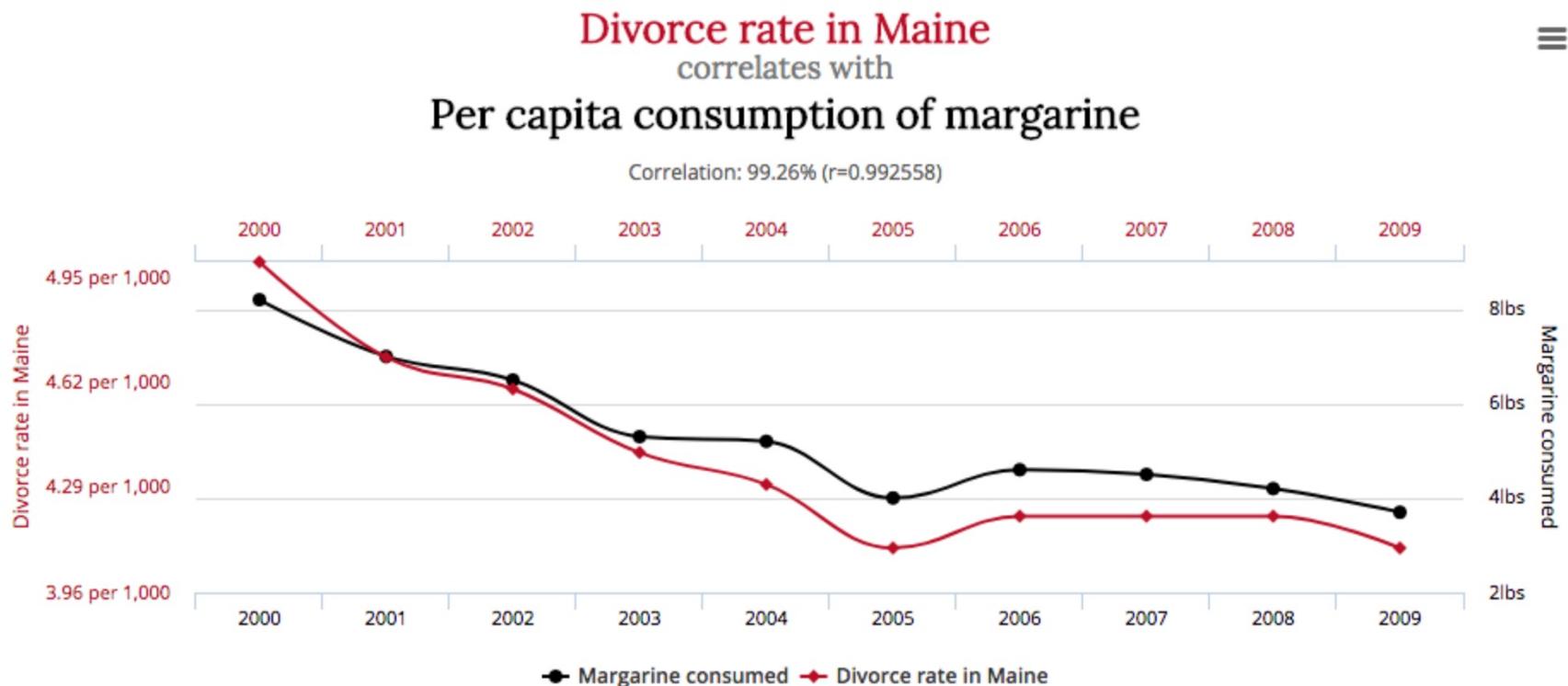
Chris Mathys



BSc Programme in Cognitive Science

Spring 2022

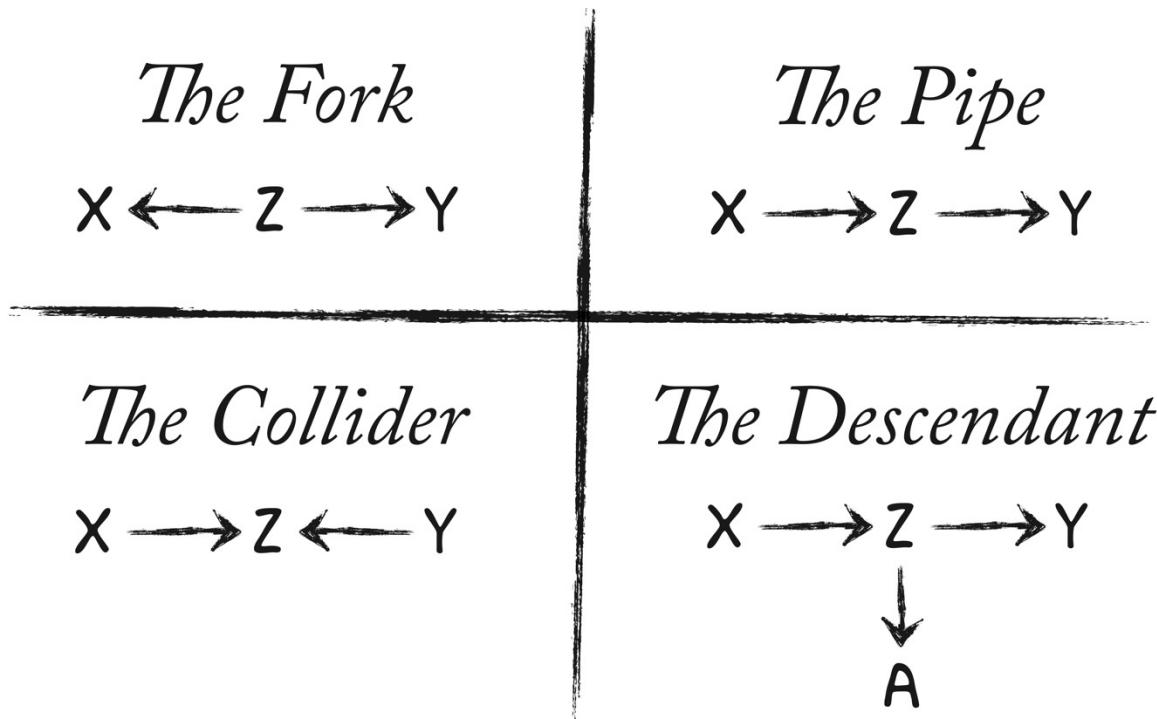
Correlation is commonplace



<http://www.tylervigen.com/spurious-correlations>

Ye Olde Causal Alchemy

The Four Elemental Confounds



The Fork

$$X \leftarrow Z \rightarrow Y$$

Z is a “confounder”

X and Y are associated

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

Share a common cause Z

Once stratified by Z , no association

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

$$X \leftarrow Z \rightarrow Y$$

```
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

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

$$Y \not\perp X$$

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

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

Y		
X	0	1
0	7	41
1	56	414

```
> 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)

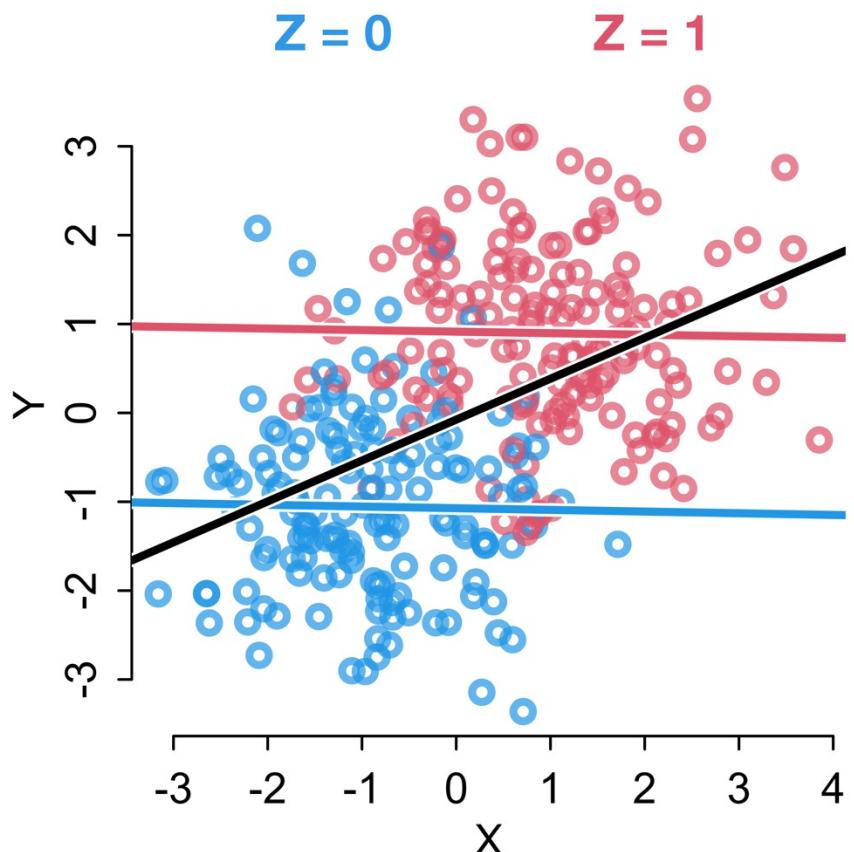
N <- 300
Z <- rbern(N)
X <- rnorm(N, 2*Z-1)
Y <- rnorm(N, 2*Z-1)

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)
```

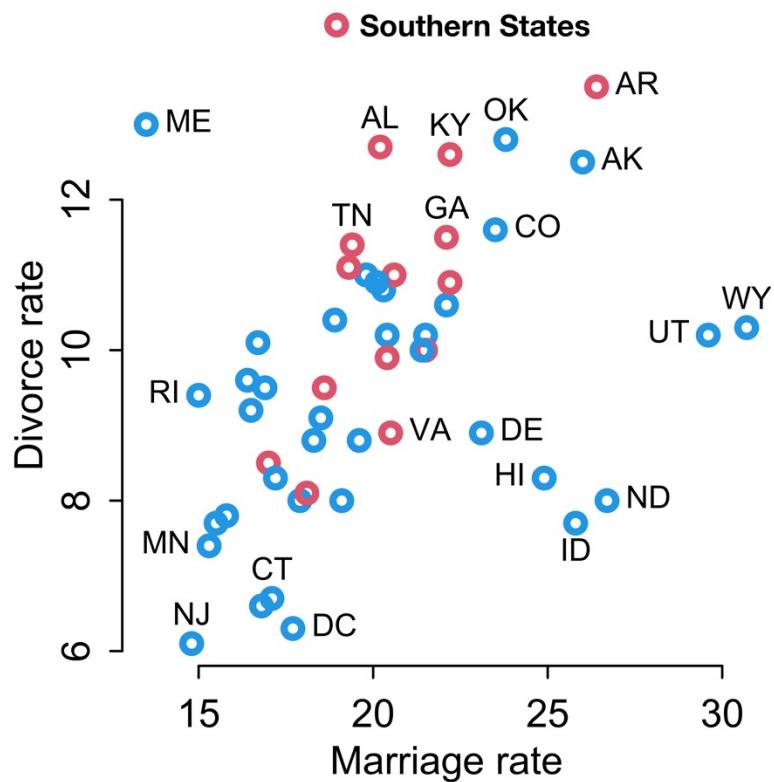


Fork Example

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

$$M \xrightarrow{?} D$$

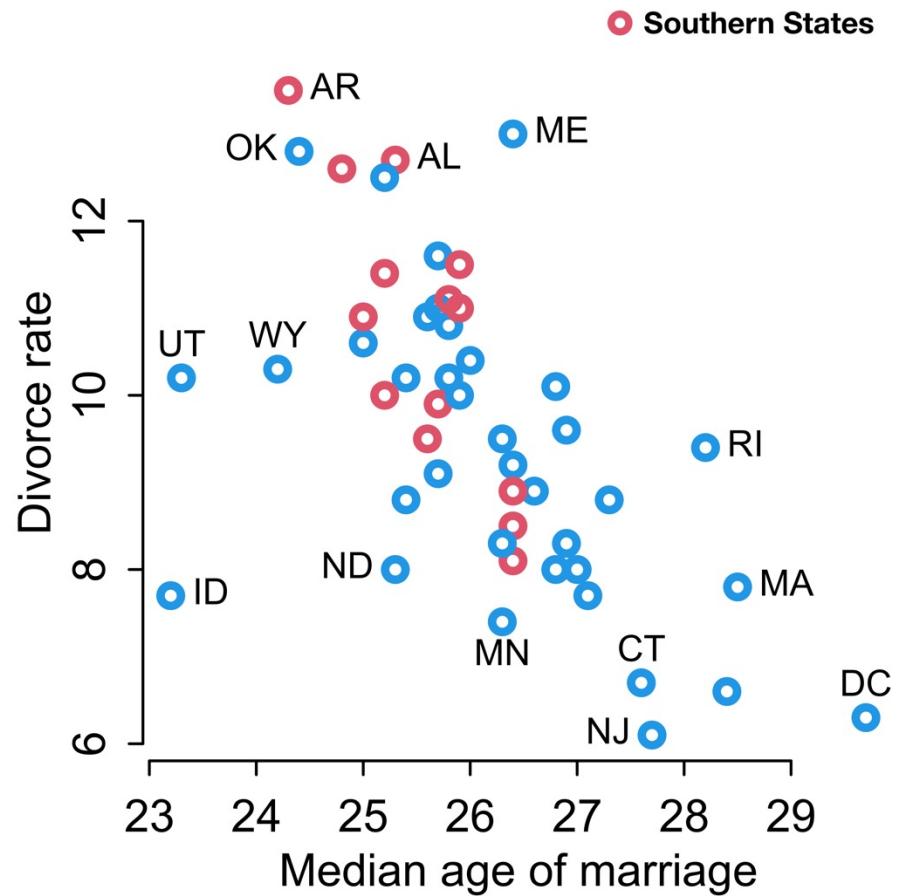
(2) Scientific model

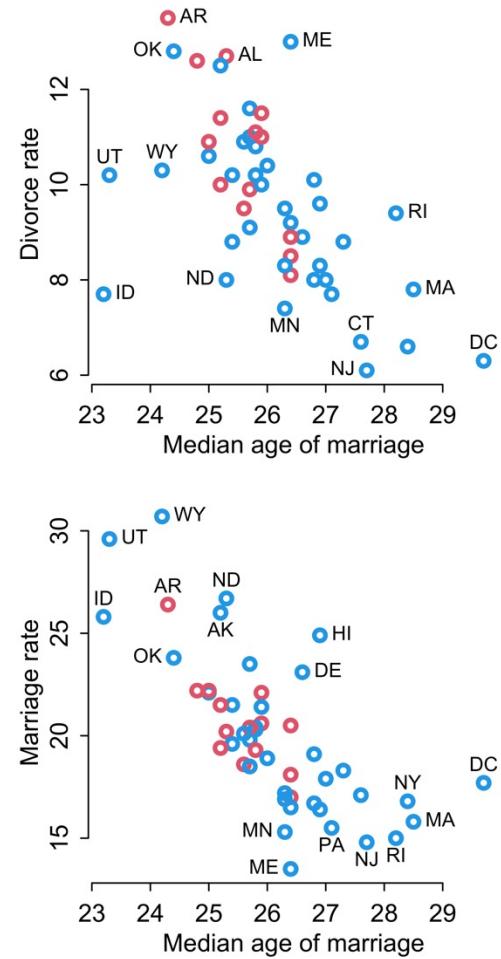
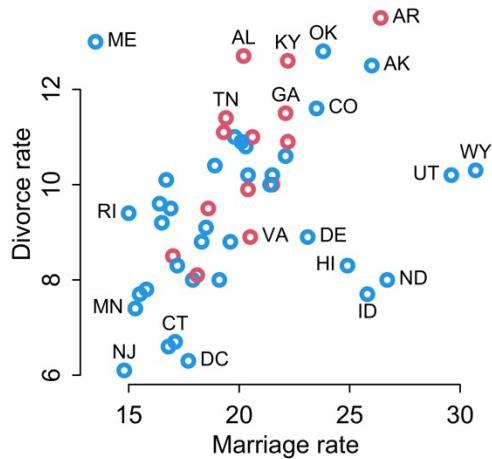
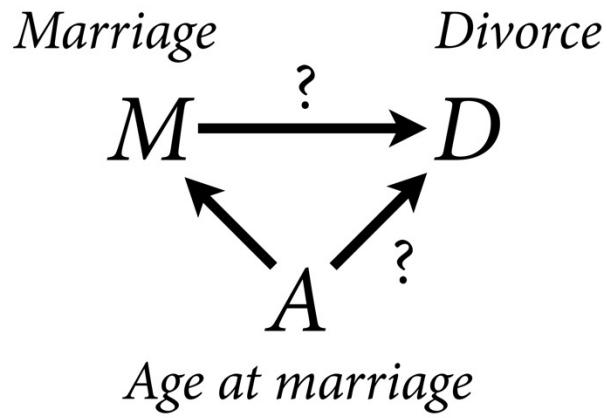
(3) Statistical model

(4) Analyze



D
Divorce
 A ?
Age at marriage





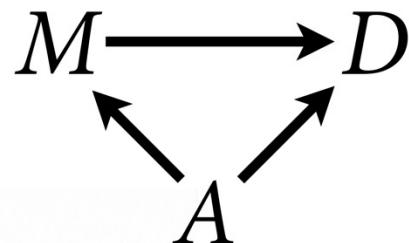
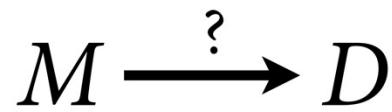
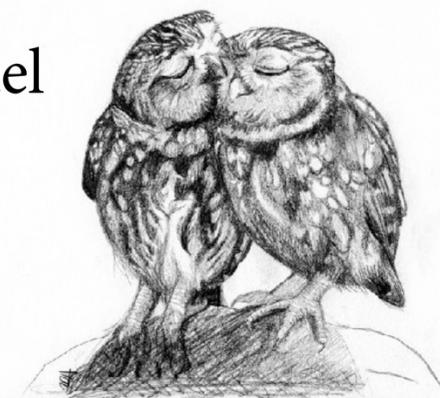
Marrying the Owl

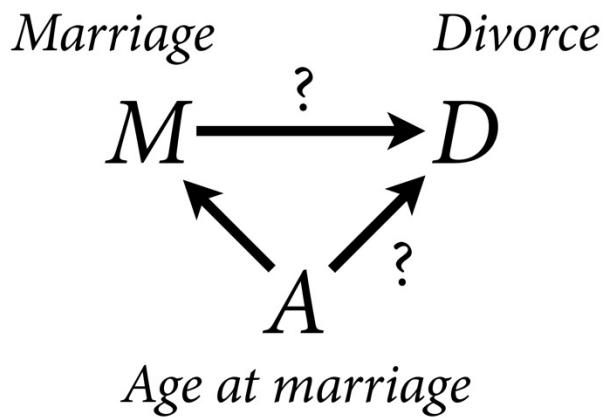
(1) Estimand: Causal effect
of **marriage** rate on **divorce**
rate

(2) Scientific model

(3) Statistical model

(4) Analyze

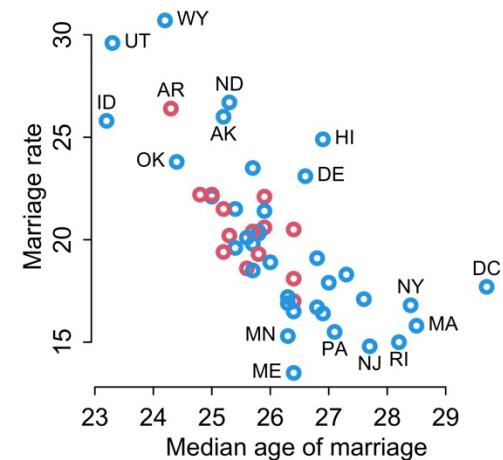
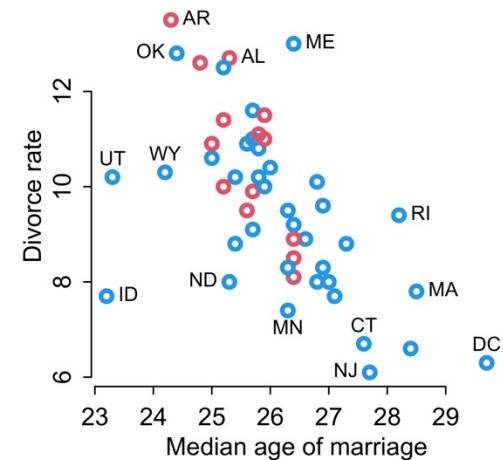
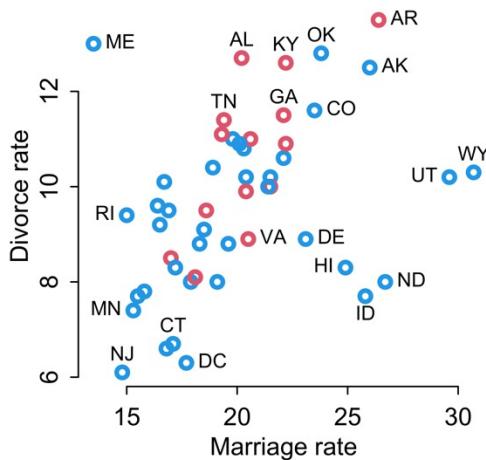




Fork: $M <- A -> D$

To estimate **direct 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 ?

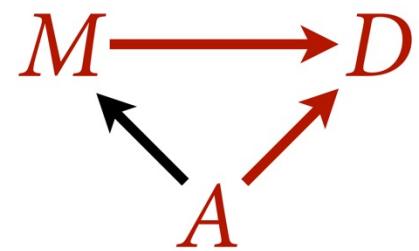
What is $D = f(A, M)$?

In a linear regression:

Text

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

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



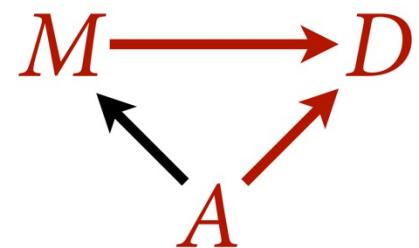
What does it mean to stratify by a continuous variable?

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

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

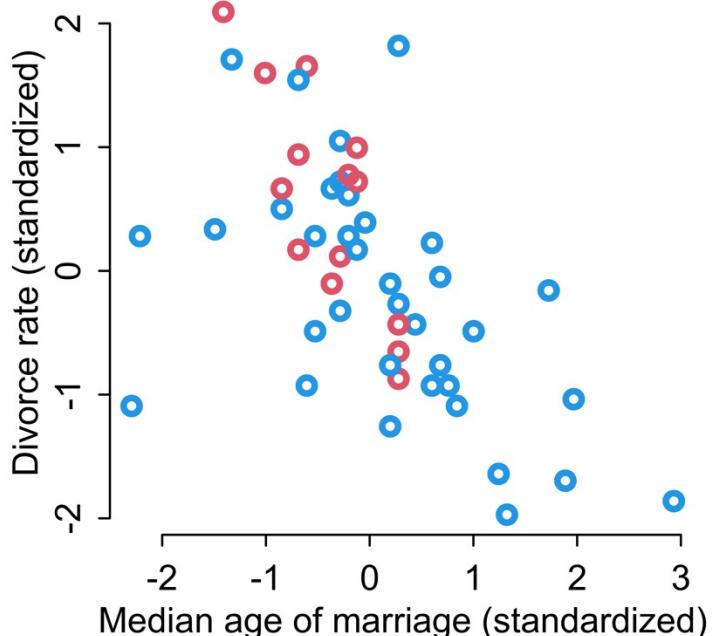


Often convenient to **standardize** variables in linear regression

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 )
}
```

These priors are crazy wide.

Once you've standardized you do not want a std dev of 10 in your priors.

Prior predictive simulation

Some default priors

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

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

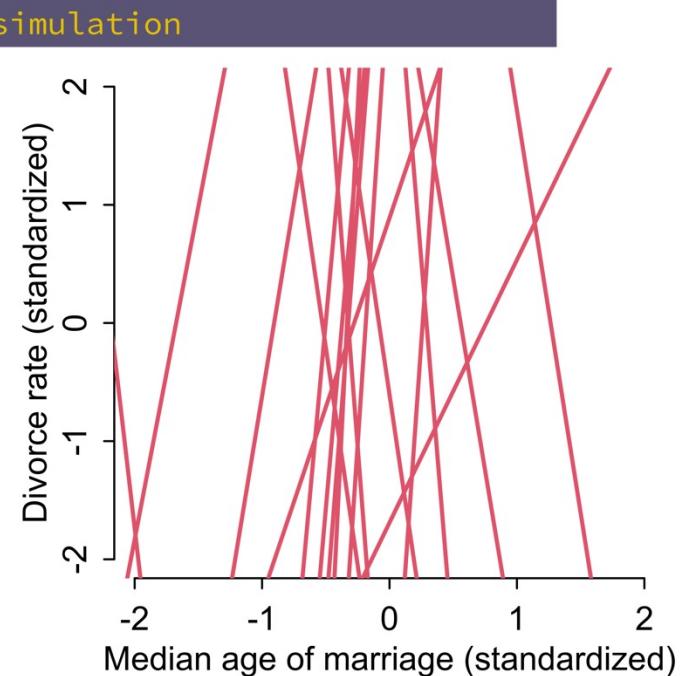
$$\alpha \sim \text{Normal}(0, 10)$$

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

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

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

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

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

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

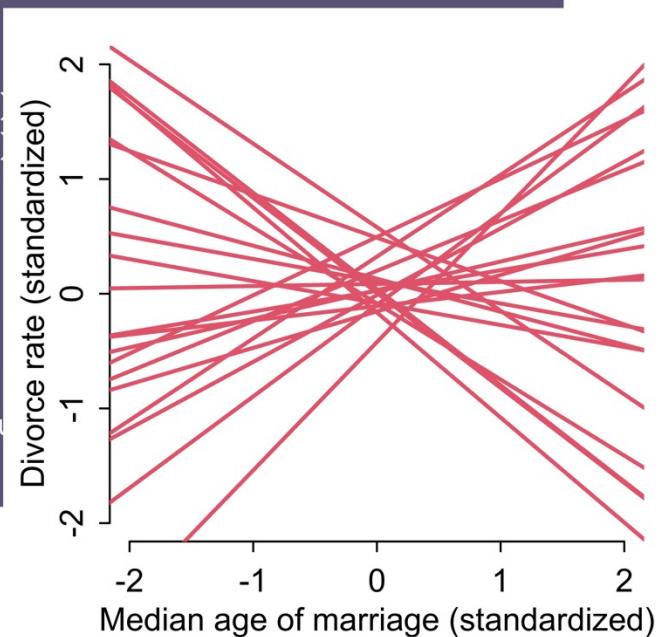
$$\alpha \sim \text{Normal}(0, 0.2)$$

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

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

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

```
# 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
xlab="Median age of
ylab="Divorce rate
Aseq <- seq(from=-3
for ( i in 1:n ) {
  mu <- a[i] + bA
  lines( Aseq , mu
}
```



Narrow priors are much better!

Marrying the Owl

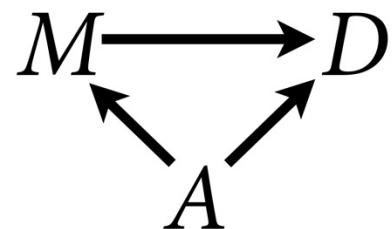
(1) Estimand: Causal effect
of **marriage** rate on **divorce**
rate

(2) Scientific model

(3) Statistical model

(4) Analyze

$$M \xrightarrow{?} D$$



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



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 )
```

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

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

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

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

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

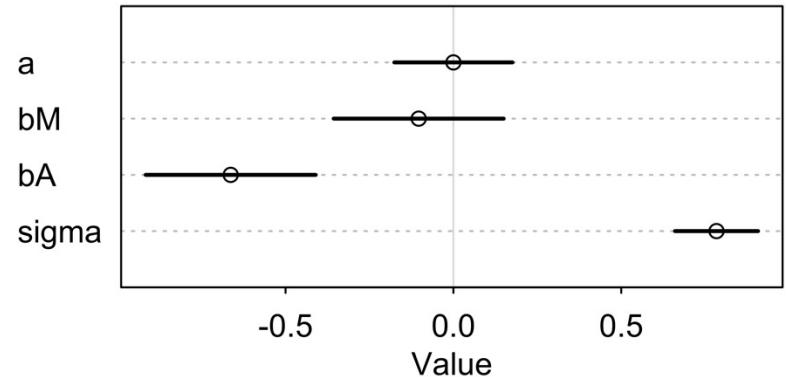
$$\sigma \sim \text{Exponential}(1)$$

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))
```



In this case, slope **bM** is estimand, but it's not always so simple

Now the data frame `d` has 100 simulated cases. Because `x_real` influences both `y` and `x_spur`, you can think of `x_spur` as another outcome of `x_real`, but one which we mistake as a potential predictor of `y`. As a result, both `x_real` and `x_spur` are correlated with `y`. You can see this in the scatterplots from `pairs(d)`. But when you include both `x` variables in a linear regression predicting `y`, the posterior mean for the association between `y` and `x_spur` will be close to zero.

5.1.5.3. Counterfactual plots. A second sort of inferential plot displays the causal implications of the model. I call these plots **COUNTERFACTUAL**, because they can be produced for any values of the predictor variables you like, even unobserved combinations like very high median age of marriage and very high marriage rate. There are no States with this combination, but in a counterfactual plot, you can ask the model for a prediction for such a State, asking questions like “What would Utah’s divorce rate be, if its median age at marriage were higher?” Used with clarity of purpose, counterfactual plots help you understand the model, as well as generate predictions for imaginary interventions and compute how much some observed outcome could be attributed to some cause.

Note that the term “counterfactual” is highly overloaded in statistics and philosophy. It hardly ever means the same thing when used by different authors. Here, I use it to indicate some computation that makes use of the structural causal model, going beyond the posterior distribution. But it could refer to questions about both the past and the future.

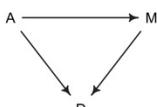
The simplest use of a counterfactual plot is to see how the outcome would change as you change one predictor at a time. If some predictor X took on a new value for one or more cases in our data, how would the outcome Y have changed? Changing just one predictor X might also change other predictors, depending upon the causal model. Suppose for example that you pay young couples to postpone marriage until they are 35 years old. Surely this will also decrease the number of couples who ever get married—some people will die before turning 35, among other reasons—decreasing the overall marriage rate. An extraordinary and evil degree of control over people would be necessary to really hold marriage rate constant while forcing everyone to marry at a later age.

So let’s see how to generate plots of model predictions that take the causal structure into account. The basic recipe is:

- (1) Pick a variable to manipulate, the intervention variable.
- (2) Define the range of values to set the intervention variable to.
- (3) For each value of the intervention variable, and for each sample in posterior, use the causal model to simulate the values of other variables, including the outcome.

In the end, you end up with a posterior distribution of counterfactual outcomes that you can plot and summarize in various ways, depending upon your goal.

Let’s see how to do this for the divorce model. Again we take this DAG as given:



To simulate from this, we need more than the DAG. We also need a set of functions that tell us how each variable is generated. For simplicity, we’ll use Gaussian distributions for each variable, just like in model `m5.3`. But model `m5.3` ignored the assumption that `A` influences

Overthinking: Simulating counterfactuals. The example in this section used `sim()` to hide the details. But simulating counterfactuals on your own is not hard. It just uses the model definition. Assume we’ve already fit model `m5.3_A`, the model that includes both causal paths $A \rightarrow D$ and $A \rightarrow M \rightarrow D$. We define a range of values that we want to assign to A :

R code
5.25

```
A_seq <- seq( from=-2 , to=2 , length.out=30 )
```

Next we need to extract the posterior samples, because we’ll simulate observations for each set of samples. Then it really is just a matter of using the model definition with the samples, as in previous examples. The model defines the distribution of M . We just convert that definition to the corresponding simulation function, which is `rnorm` in this case:

R code
5.26

```
post <- extract.samples( m5.3_A )
M_sim <- with( post , sapply( 1:30 ,
  function(i) rnorm( 1e3 , aM + bAM*A_seq[i] , sigma_M ) ) )
```

I used the `with` function, which saves us having to type `post$` in front of every parameter name. The linear model inside `rnorm` comes right out of the model definition. This produces a matrix of values, with samples in rows and cases corresponding to the values in `A_seq` in the columns. Now that we have simulated values for M , we can simulate D too:

R code
5.27

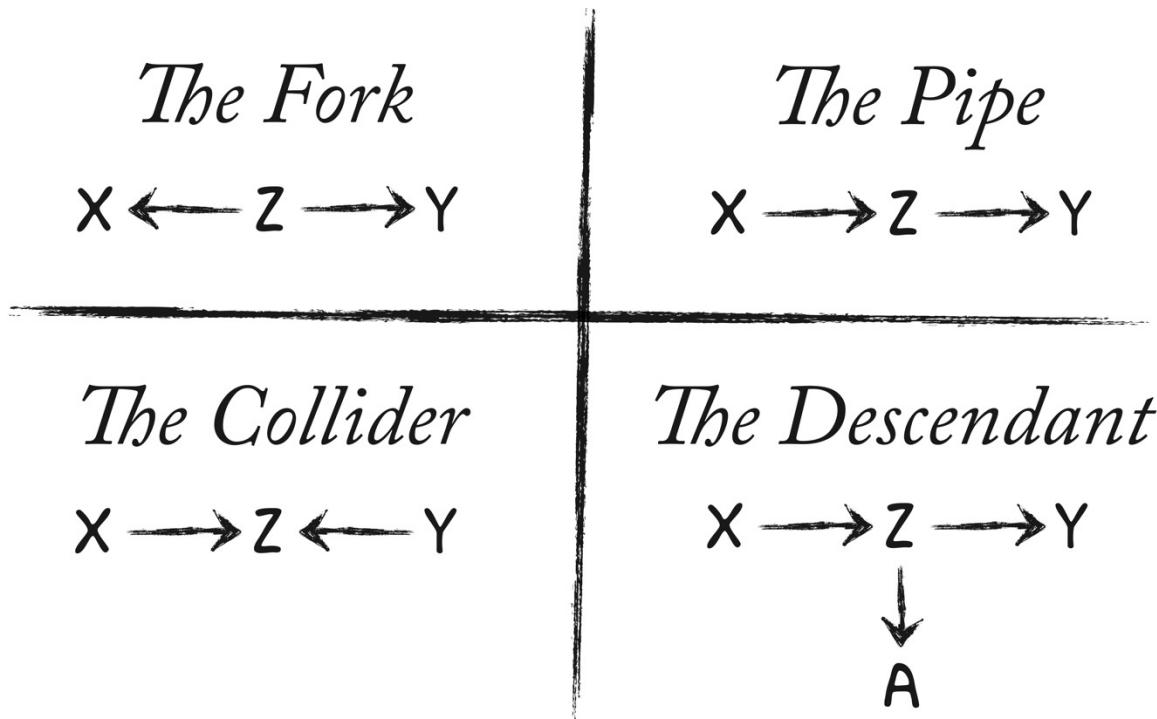
```
D_sim <- with( post , sapply( 1:30 ,
  function(i) rnorm( 1e3 , a + bA*A_seq[i] + bM*M_sim[,i] , sigma ) ) )
```

If you plot `A_seq` against the column means of `D_sim`, you’ll see the same result as before. In complex models, there might be many more variables to simulate. But the basic procedure is the same.

Simulating causal effects: Page 140–144

Ye Olde Causal Alchemy

The Four Elemental Confounds



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

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

$X \rightarrow Z \rightarrow Y$

```
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 )
```

X	Y	
	0	1
0	430	87
1	93	390

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

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

Z	Y	
	0	1
0	422	39
1	53	5

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

X	Y	
	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)

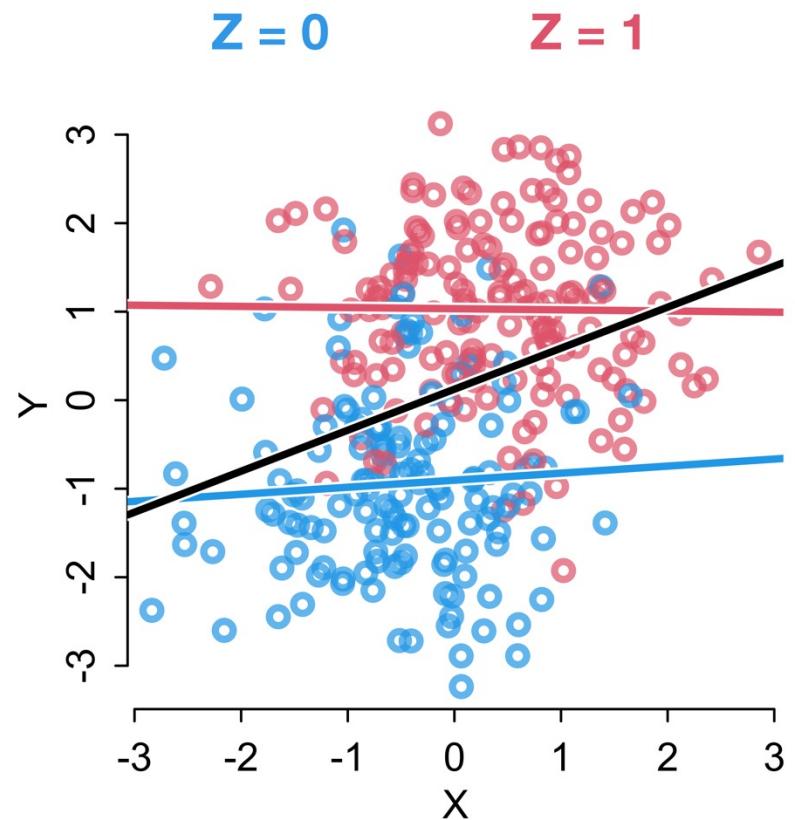
N <- 300
X <- rnorm(N)
Z <- rbern(N,inv_logit(X))
Y <- rnorm(N,(2*Z-1))

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

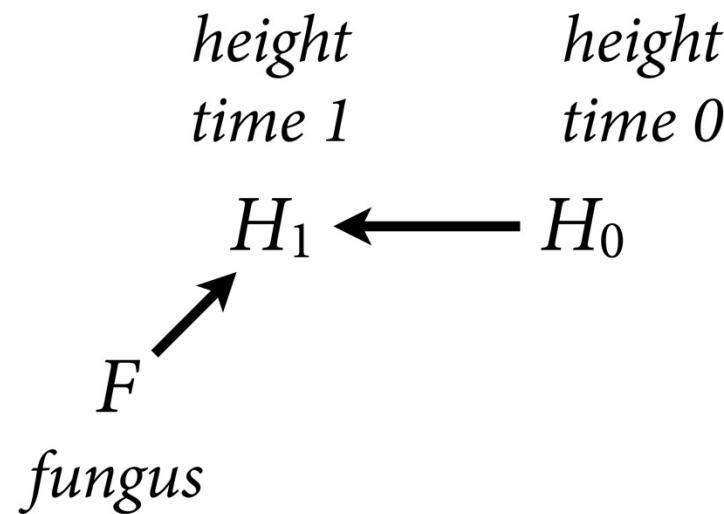
Half treated with anti-fungal

Measure growth and fungus

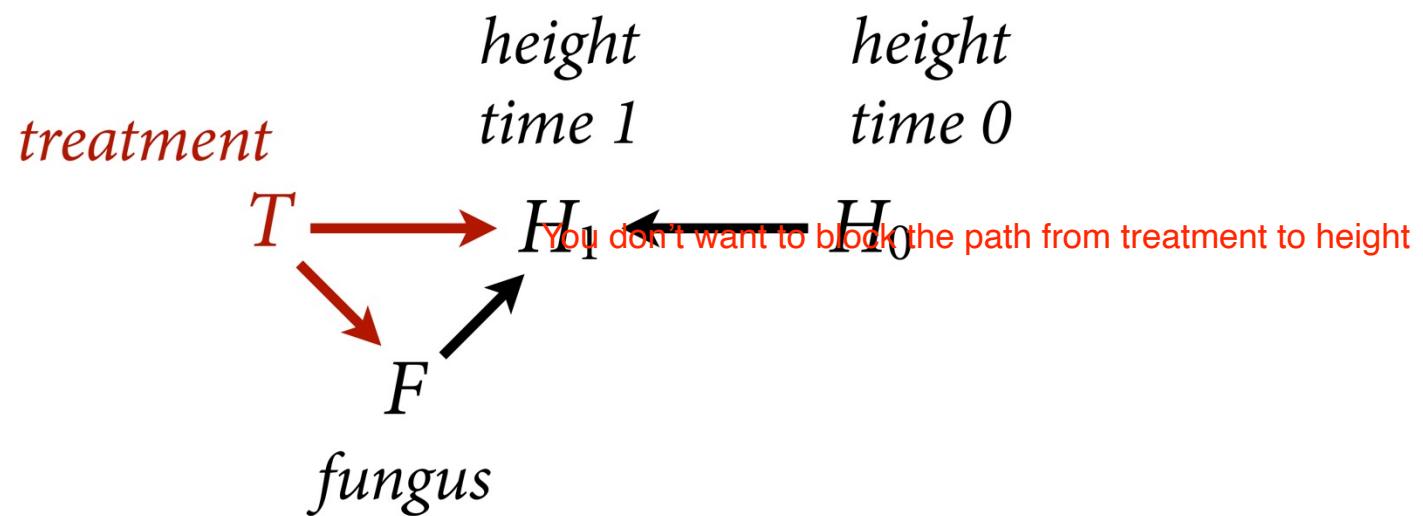
Estimand: Causal effect of treatment on plant growth



Scientific model



Scientific model



Statistical model

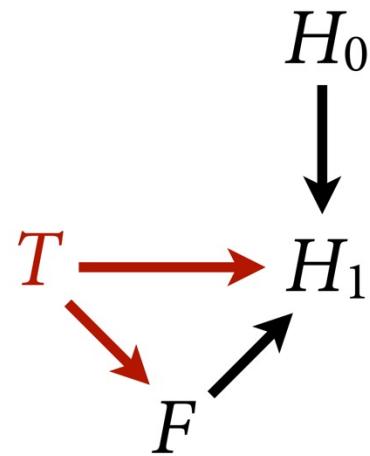
Estimand: Total causal effect of T

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**

Might conclude that treatment doesn't work when it actually does

Consequences of treatment should not usually be included in your statistical model (do include in scientific model!)

Doing experiments is no protection against bad causal inference

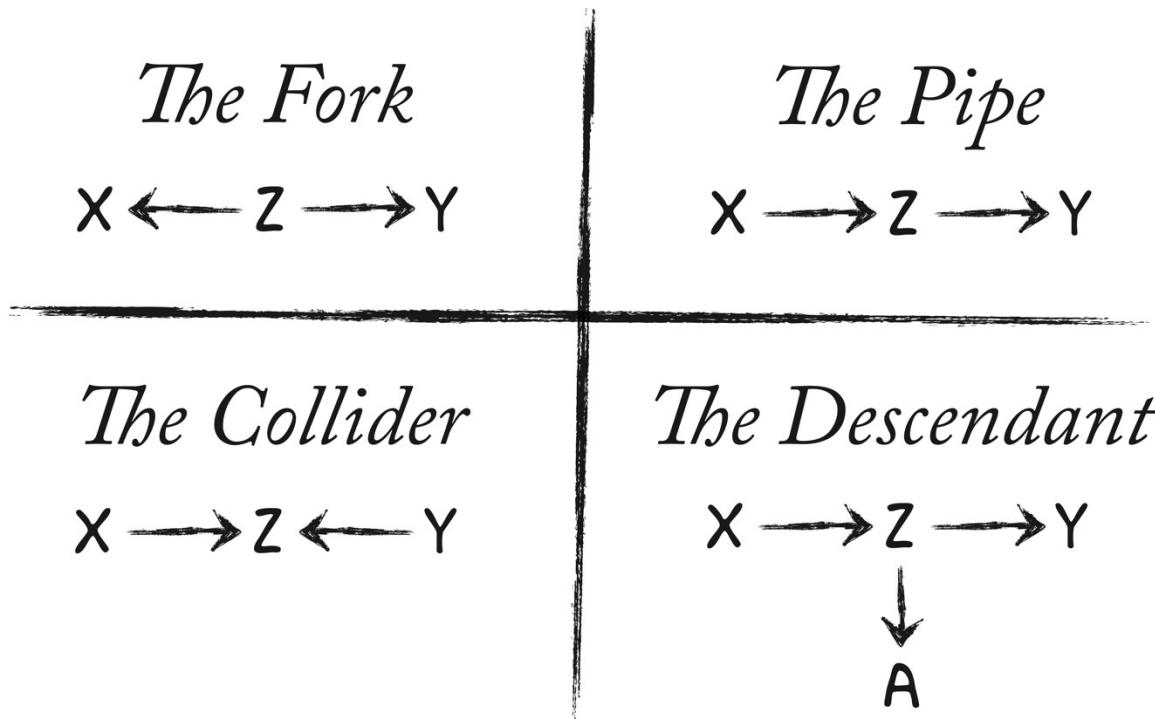
TABLE 1 Posttreatment Conditioning in Experimental Studies

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”

Ye Olde Causal Alchemy

The Four Elemental Confounds



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

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

$$X \rightarrow Z \leftarrow Y$$

```
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
```

$$X \rightarrow Z \leftarrow Y$$

```
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
```

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

Z = 1	$Y \not\perp X Z$
	Y
X	0 1
0	43 217
1	218 242

```
> 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))

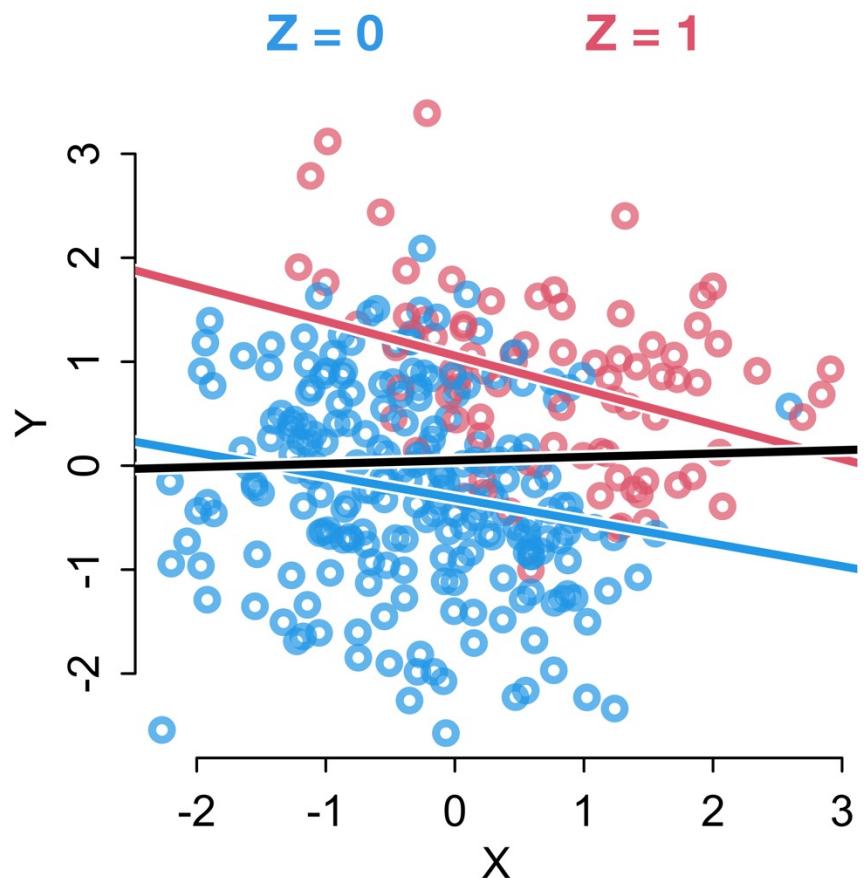
plot( X , Y , col=cols[Z+1] , lwd=3 )

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)

```



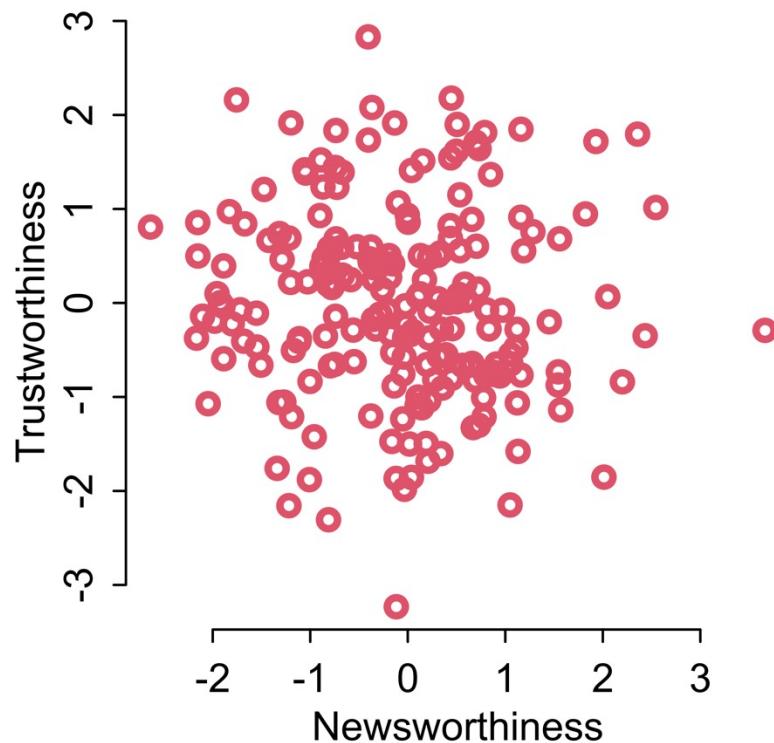
Collider example

Some biases arise from selection

Suppose: 200 grant applications

Each scored on newsworthiness
and trustworthiness

No association in population



Collider example

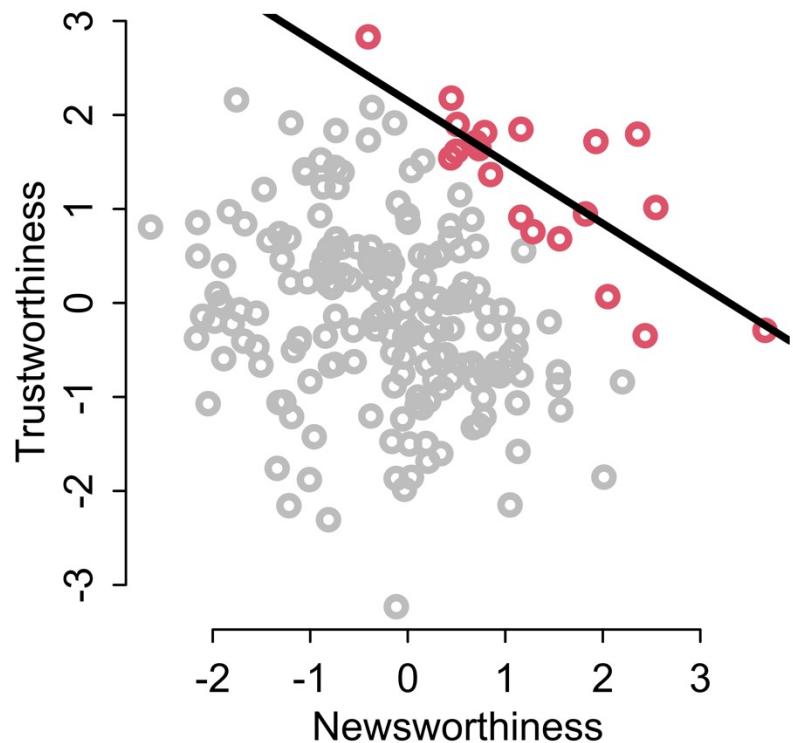
Some biases arise from selection

Suppose: 200 grant applications

Each scored on newsworthiness
and trustworthiness

No association in population

Strong association after selection



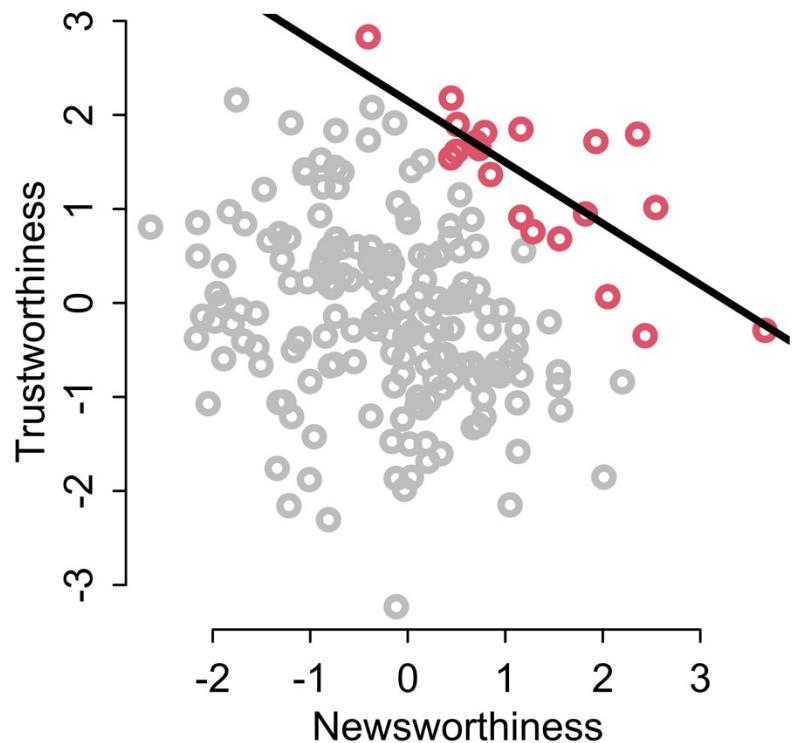
Collider example

$$N \rightarrow A \leftarrow T$$

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

Few grants are high in both

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



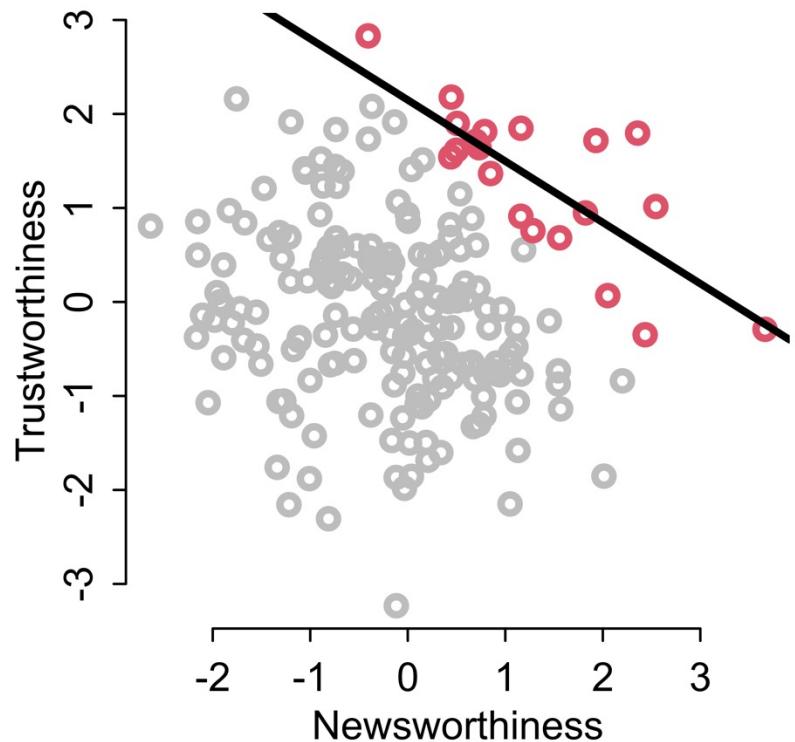
Collider example

$$N \rightarrow A \leftarrow T$$

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

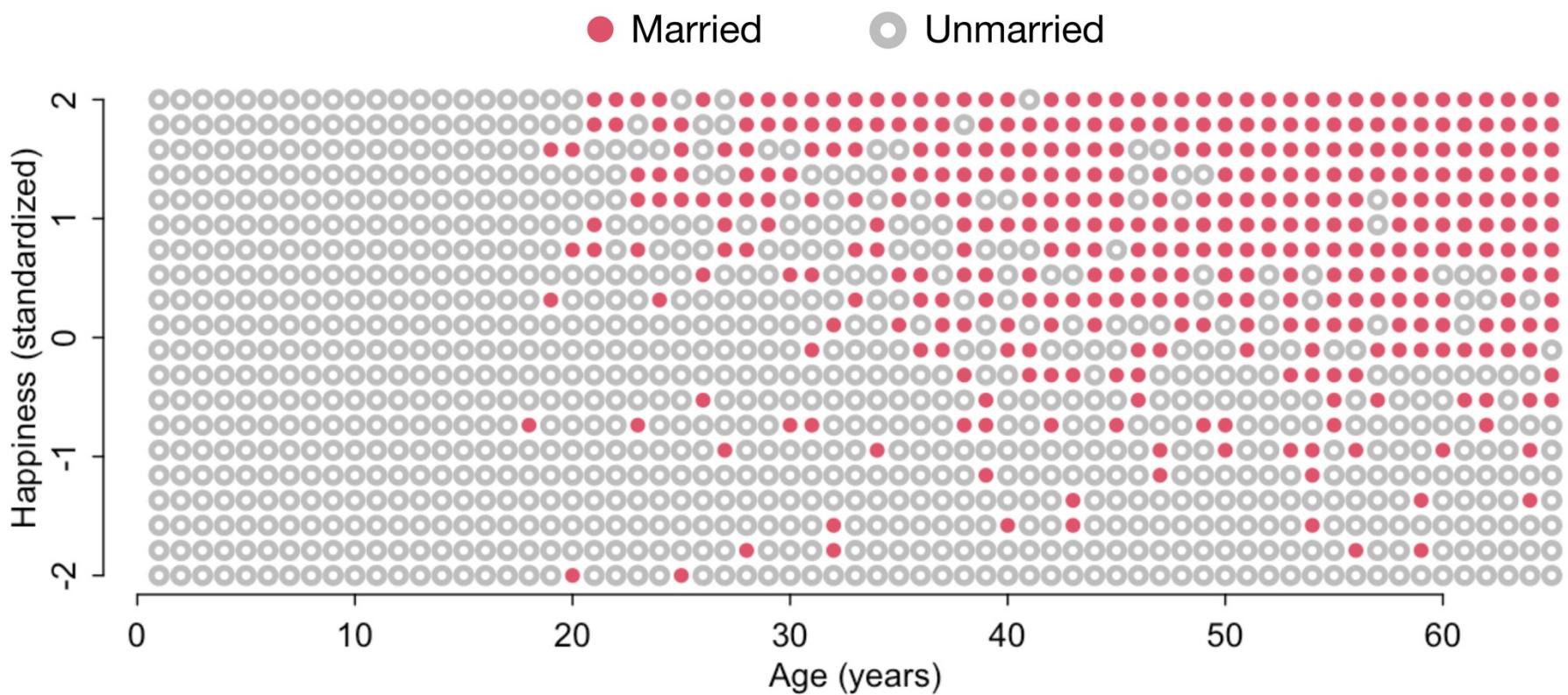
Possible confound: Marital status

Suppose age has zero influence on happiness

But that both age and happiness influence marital status

$$H \rightarrow M \leftarrow A$$

*Happiness
Married
Age*



Page 177

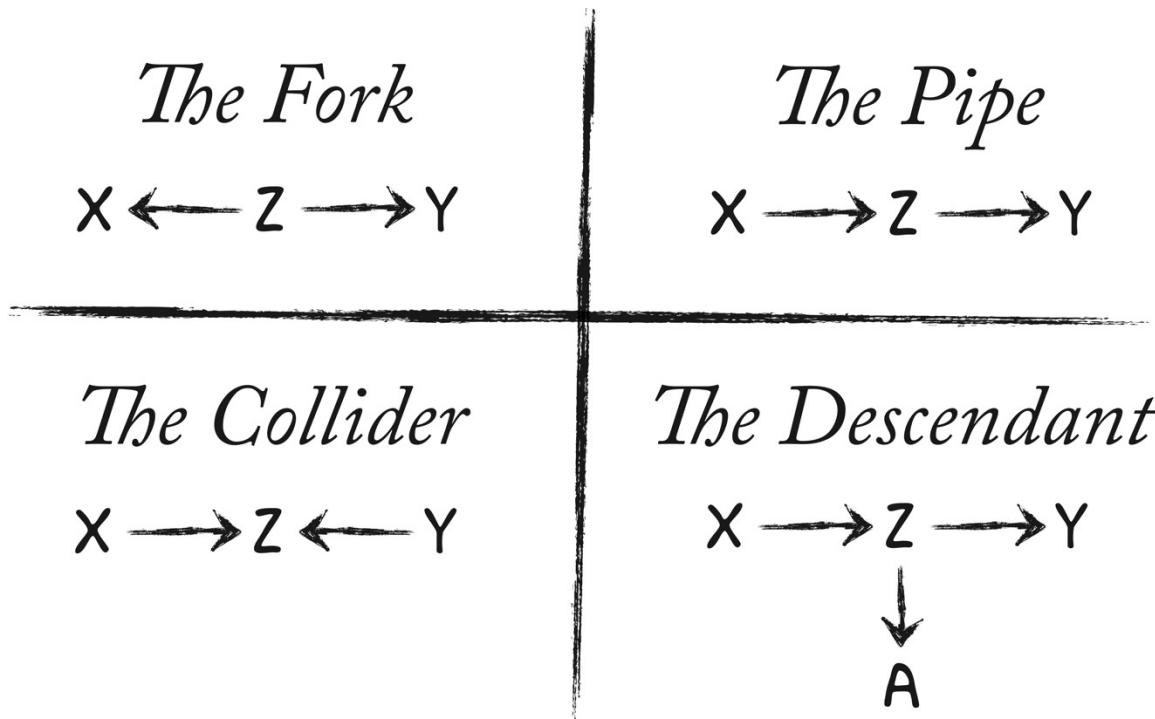
Stratified by marital status, negative association between age and happiness



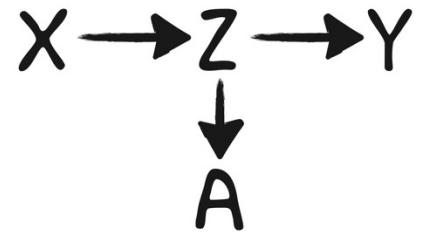
Full workflow starting on page 176

Ye Olde Causal Alchemy

The Four Elemental Confounds



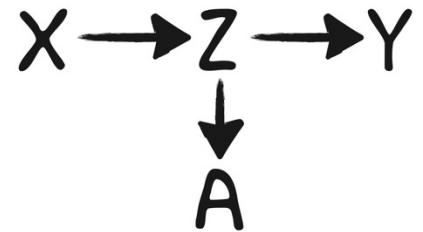
The Descendant



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

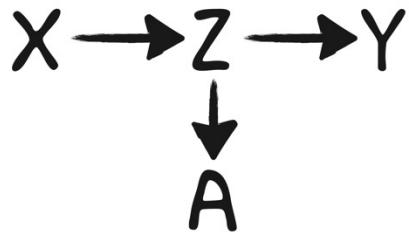
$Y \not\perp\!\!\!\perp X$

A holds information about Z

if strong enough

Once stratified by A , X and Y less associated

$Y \perp\!\!\!\perp X | A$



```

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

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

```

> cor(X,Y)
[1] 0.61

```

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

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

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

```

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

```

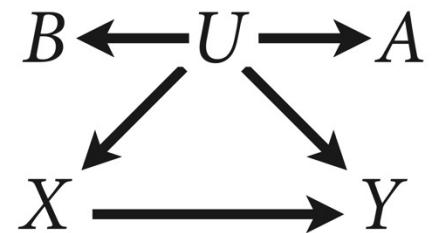
Descendants are everywhere

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 ?

