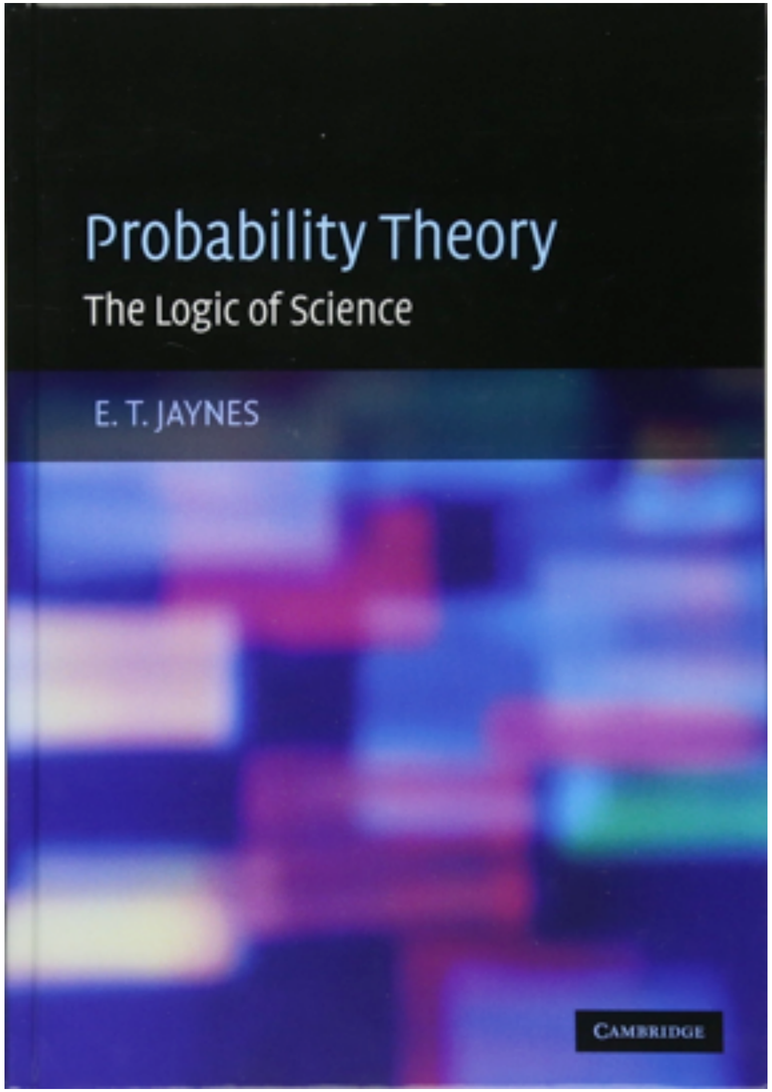
Why normal?

- Ontological perspective
  - Processes which add fluctuations result in dampening
  - Damped fluctuations end up Gaussian
  - No information left, except mean and variance
  - Can't infer process from distribution!
- Epistemological perspective
  - Know only *mean* and *variance*
  - Then least surprising and most conservative (*maximum entropy*) distribution is Gaussian
  - Nature likes maximum entropy distributions



Language for modeling

Give them definitions:

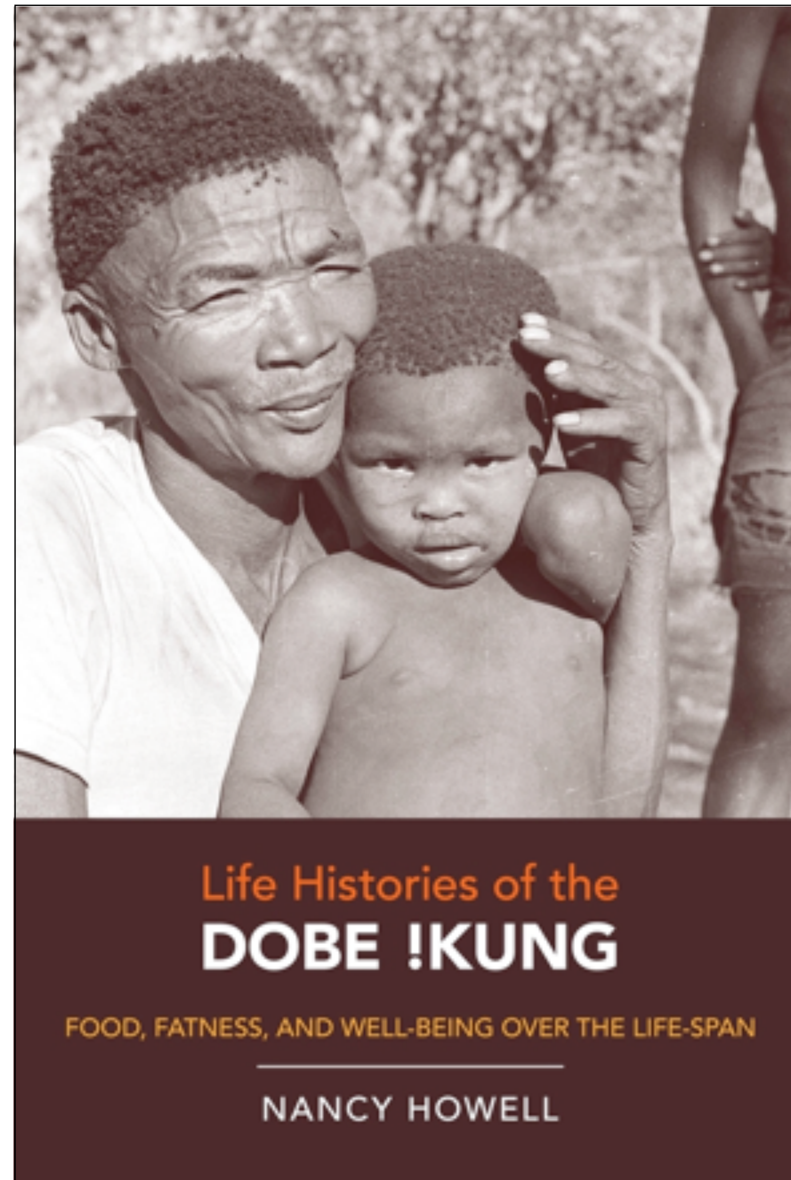
$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \beta x_i \\
\beta &\sim \text{Normal}(0, 10) \\
\sigma &\sim \text{Exponential}(1) \\
x_i &\sim \text{Normal}(0, 1)
\end{aligned}
$$

Some data: Kalahari foragers

```
library(rethinking)
data(Howell1)
d <- Howell1
```

```
precis( d )
```

```
'data.frame': 544 obs. of 4 variables:
         mean    sd  5.5%  94.5%     histogram
height 138.26 27.60 81.11 165.74
weight  35.61 14.72  9.36  54.50
age     29.34 20.75  1.00  66.13
male     0.47  0.50  0.00   1.00
```



Adding a predictor variable

- Use a linear model of the mean, *mu*:

$$
\begin{aligned}
h_i &\sim \text{Normal}(\mu_i, \sigma) && \text{[likelihood]} \\
\mu_i &= \alpha + \beta(x_i - \bar{x}) && \text{[linear model]} \\
\alpha &\sim \text{Normal}(178, 20) && [\alpha \text{ prior}] \\
\beta &\sim \text{Normal}(0, 10) && [\beta \text{ prior}] \\
\sigma &\sim \text{Uniform}(0, 50) && [\sigma \text{ prior}]
\end{aligned}
$$

Adding a predictor variable

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

mean on row i

$$\mu_i = \alpha + \beta(x_i - \bar{x})$$

mean weight

mean when $x_i = 0$

"intercept"

weight on row i

change in mean, per unit change x_i

"slope"

Prior predictive distribution

```
set.seed(2971)
N <- 100                          # 100 lines
a <- rnorm( N , 178 , 20 )
b <- rnorm( N , 0 , 10 )
```

R code 4.38

$$\alpha \sim \text{Normal}(178, 20)$$
$$\beta \sim \text{Normal}(0, 10)$$

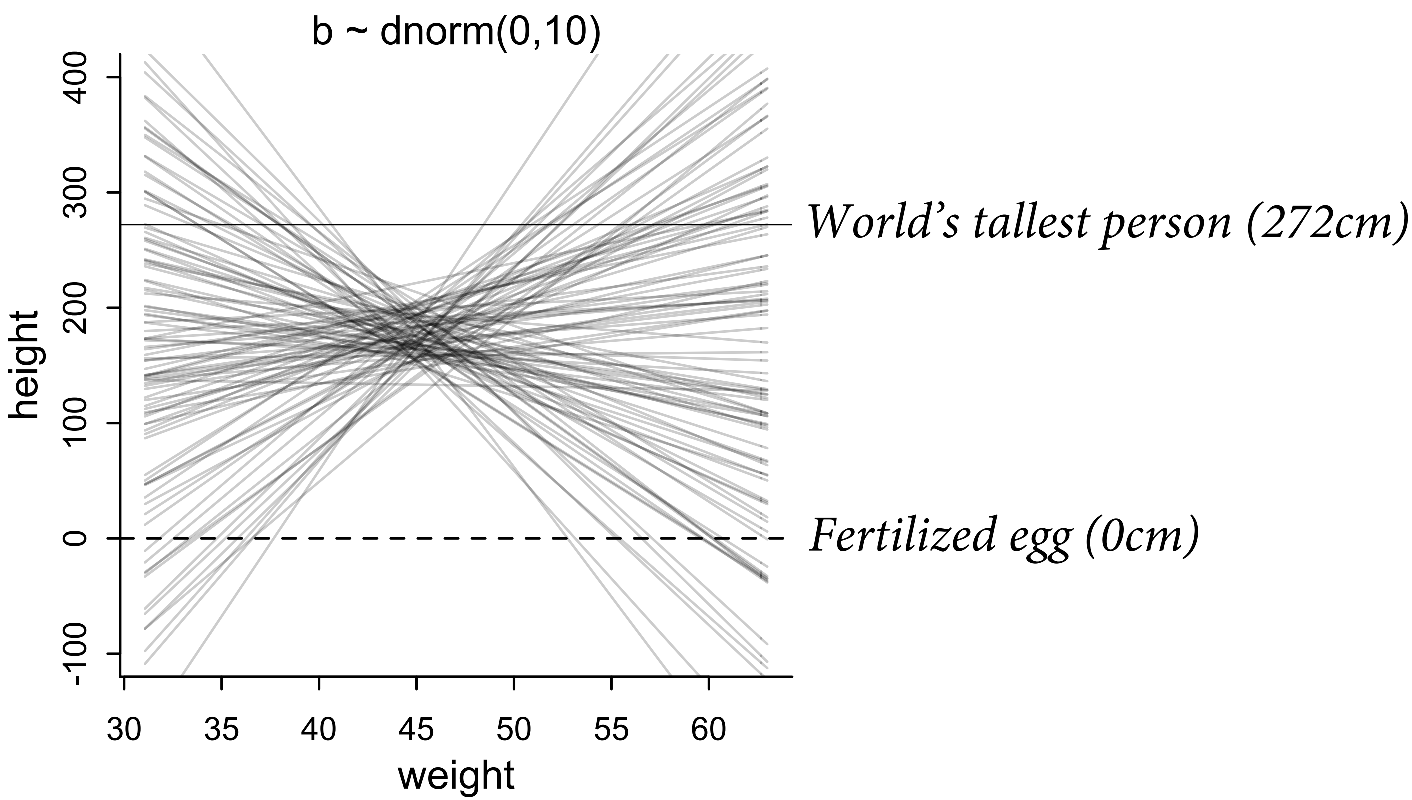


Figure 4.5

Prior predictive distribution

R code
4.41

```
set.seed(2971)
N <- 100                    # 100 lines
a <- rnorm( N , 178 , 20 )
b <- rlnorm( N , 0 , 1 )
```

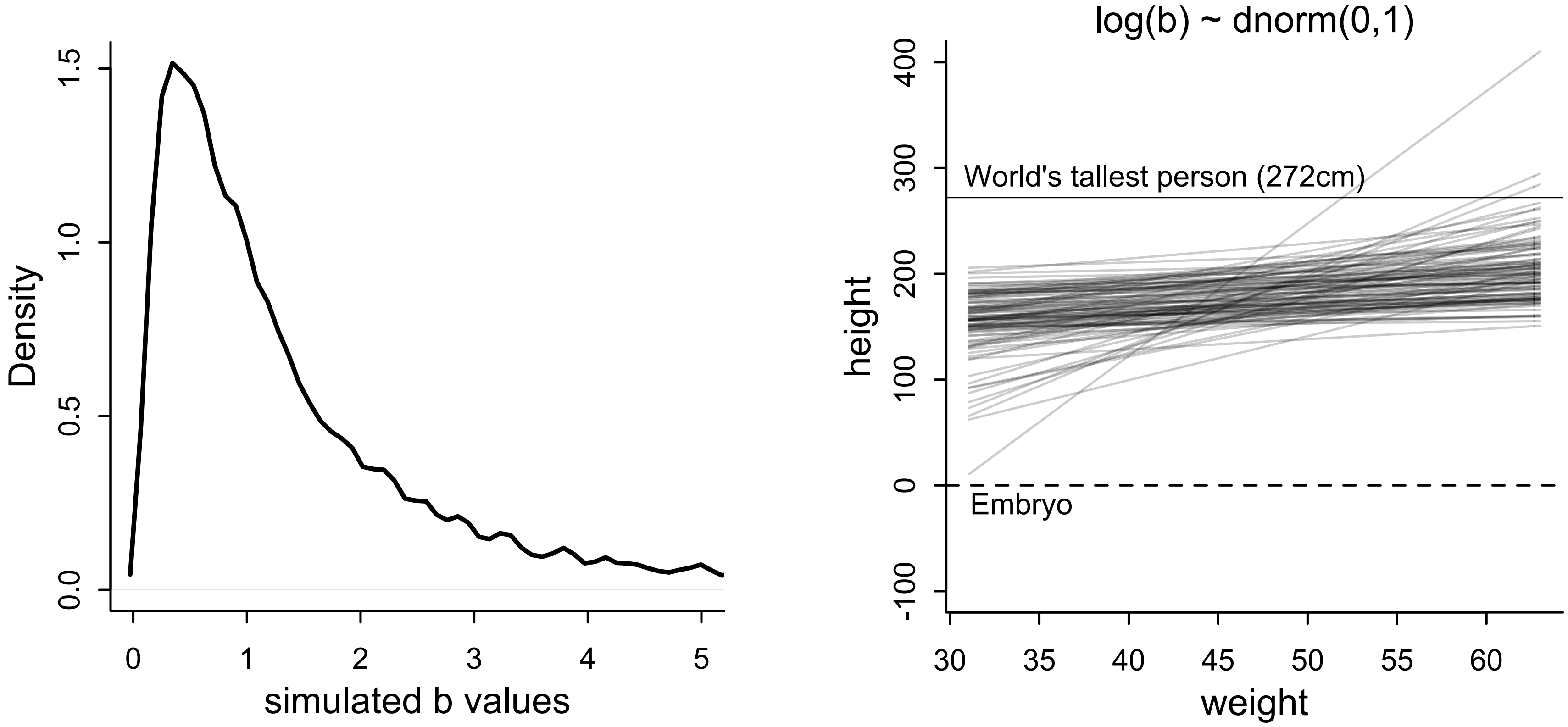
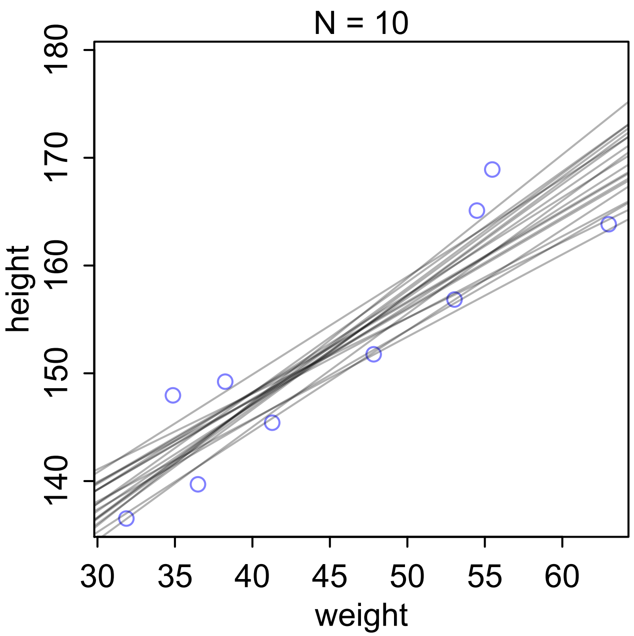


Figure 4.5

Sampling from the posterior

- Want to get uncertainty onto that graph
- Again, sample from posterior
  1. Use mean and standard deviation to approximate posterior
  2. Sample from *multivariate normal* distribution of parameters
  3. Use samples to generate predictions that "integrate over" the uncertainty

Posterior is full of lines



R code 4.47

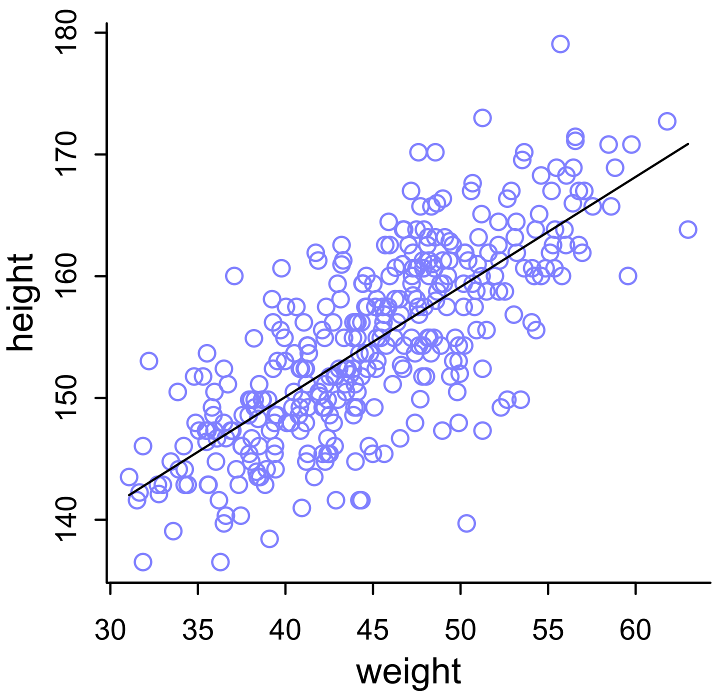
```
post <- extract.samples( m4.3 )
post[1:5,]
```

```
         a         b    sigma
1 154.5505 0.9222372 5.188631
2 154.4965 0.9286227 5.278370
3 154.4794 0.9490329 4.937513
4 155.2289 0.9252048 4.869807
5 154.9545 0.8192535 5.063672
```

Figure 4.7

Showing Uncertainty

- Want to get uncertainty onto that graph
- Again, sample from posterior
  1. Use mean and standard deviation to approximate posterior
  2. Sample from *multivariate normal* distribution of parameters
  3. Use samples to generate predictions that *integrate over* the uncertainty



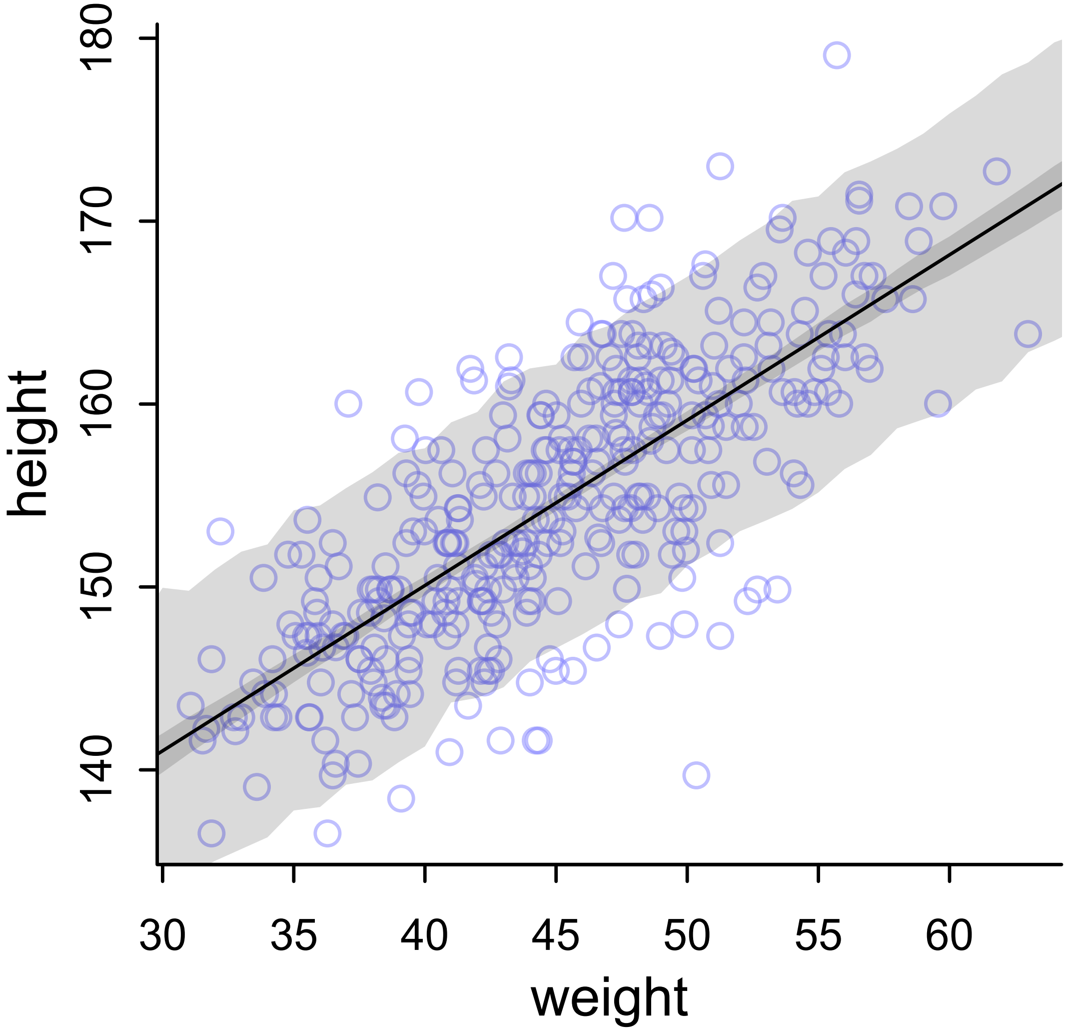
How `link` works

- Sample from posterior
- Define series of predictor (weight) values
- For each predictor value
  - For each sample from posterior
    - Compute mu: a + b*(weight – xbar)
- Summarize

R code
4.58

```
post <- extract.samples(m4.3)
mu.link <- function(weight) post$a + post$b*( weight - xbar )
weight.seq <- seq( from=25 , to=70 , by=1 )
mu <- sapply( weight.seq , mu.link )
mu.mean <- apply( mu , 2 , mean )
mu.HPDI <- apply( mu , 2 , HPDI , prob=0.89 )
```

89% prediction interval



Nothing special about 95%

Try 50%, 80%, 99%

Interested in *shape*,
not *boundaries*

Figure 4.10

Curves From Lines

- "Linear" models can make curves
- Polynomial regression
  - Common
  - Badly behaved
- Splines
  - Very flexible
  - Highly geocentric

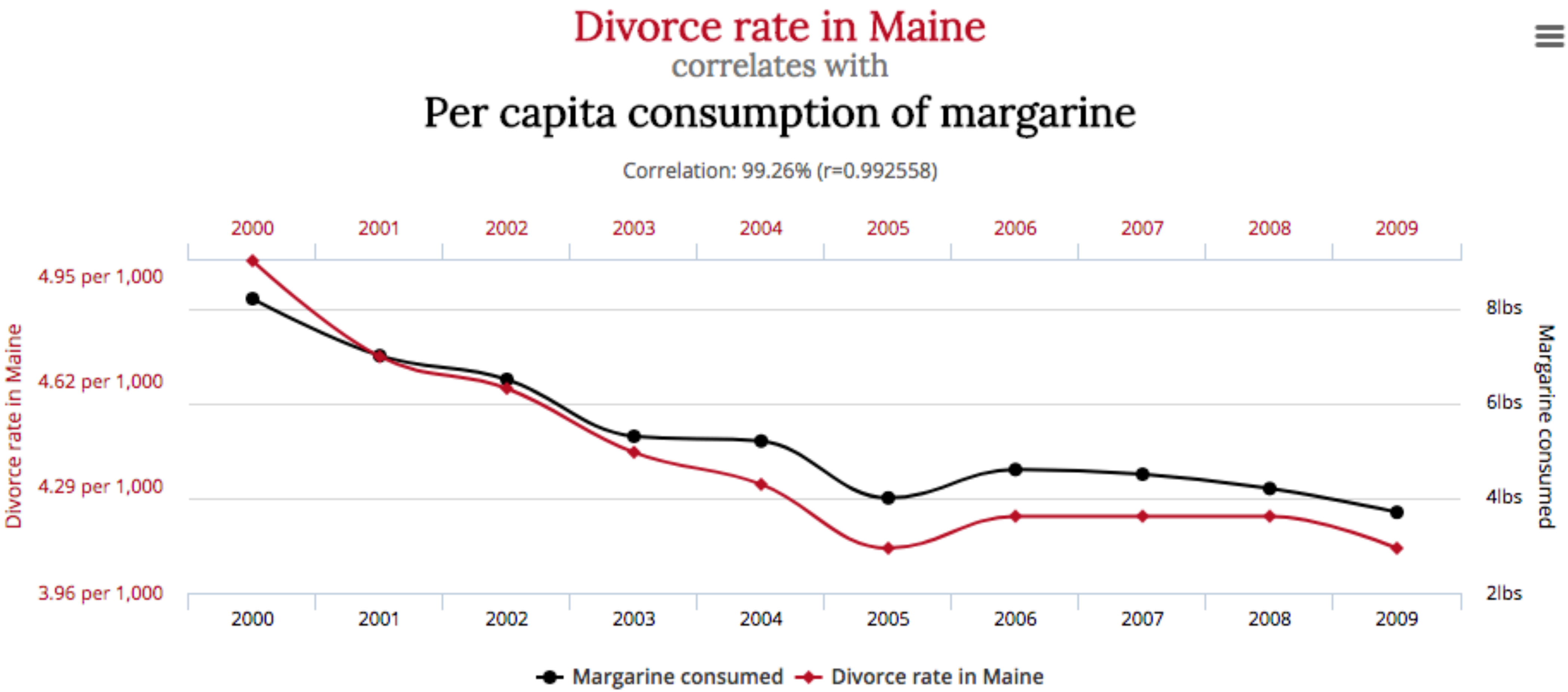
Going Local — B-Splines

- B-Splines are just linear models, but with some weird synthetic variables:

$$\mu_i = \alpha + w_1 B_{i,1} + w_2 B_{i,2} + w_3 B_{i,3} + \ldots$$

- Weights w are like slopes
- Basis functions B are synthetic variables
  - In spirit like a squared or cubed terms
  - But observed data not used to build B
  - B values turn on weights in different regions of x variable

Correlation is commonplace

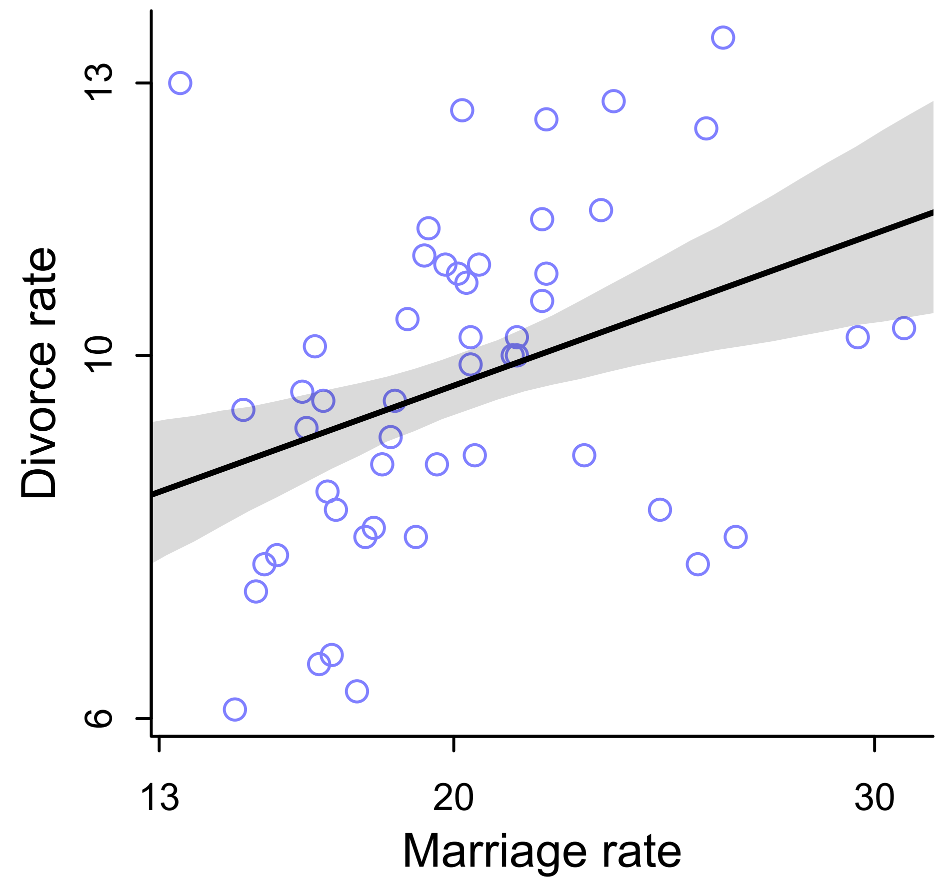


http://www.tylervigen.com/spurious-correlations

Spurious association

- Correlation does not imply causation
- Causation does not imply correlation
- Causation implies conditional correlation
- Need more than just models

- Q: Does marriage cause divorce?

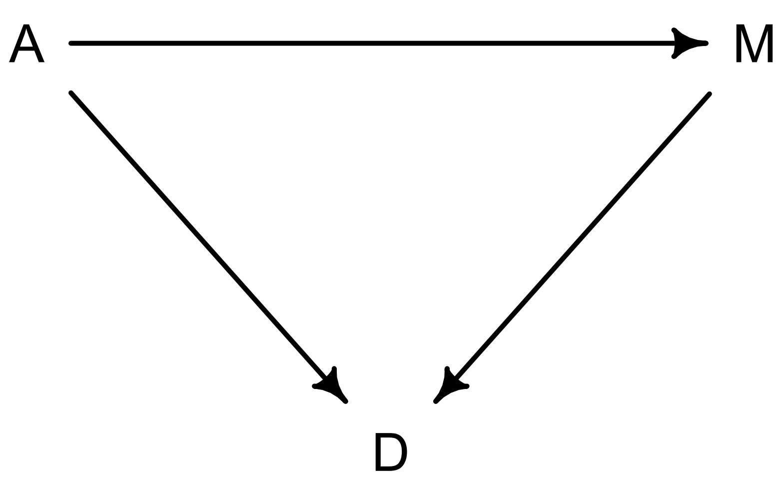


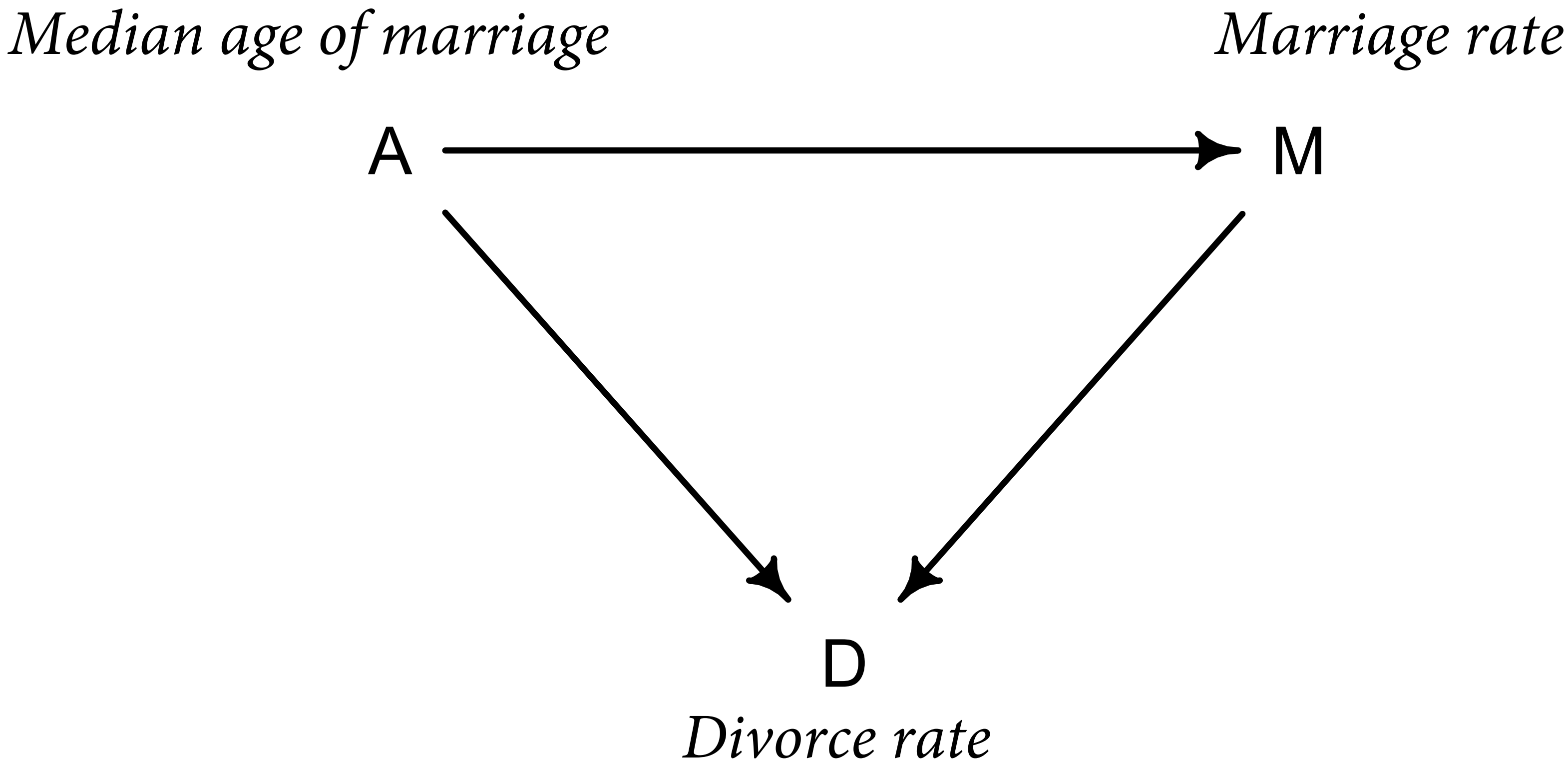
Multiple causes of divorce

- Want to know: *what is value of a predictor, once we know the other predictors?*
    - What is value of knowing marriage rate, once we already know median age at marriage?
    - What is value of knowing median age marriage, once we know marriage rate?

They're good DAGs, Brent

- Directed Acyclic Graphs — tools for causal models
  - Directed: Arrows
  - Acyclic: Arrows don't make loops
  - Graphs: Nodes and edges
- Unlike statistical model, has causal implications





Implications:

(1) M is a function of A

(2) D is a function of A and M

(3) The total causal effect of A has two *paths*:

(a) A –> M –> D

(b) A –> D

divorce rate

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

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

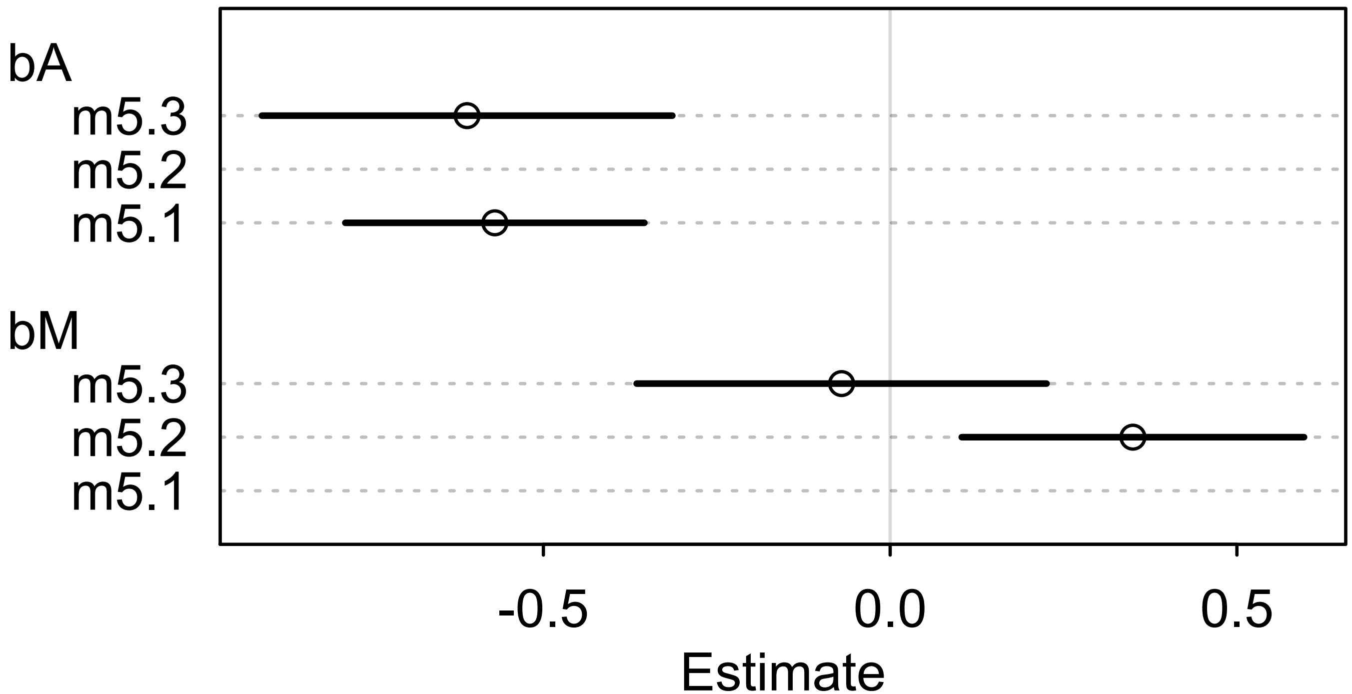
"slope" for marriage rate

marriage rate

"slope" for median age marriage

median age marriage

$$
\begin{aligned}
D_i &\sim \text{Normal}(\mu_i, \sigma) && \text{[probability of data]} \\
\mu_i &= \alpha + \beta_M M_i + \beta_A A_i && \text{[linear model]} \\
\alpha &\sim \text{Normal}(0, 0.2) && \text{[prior for } \alpha\text{]} \\
\beta_M &\sim \text{Normal}(0, 0.5) && \text{[prior for } \beta_M\text{]} \\
\beta_A &\sim \text{Normal}(0, 0.5) && \text{[prior for } \beta_A\text{]} \\
\sigma &\sim \text{Exponential}(1) && \text{[prior for } \sigma\text{]}
\end{aligned}
$$



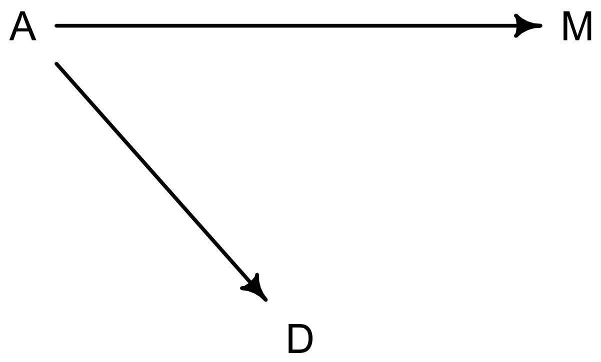
m5.1: age of marriage only D ~ A

m5.2: marriage rate only D ~ M

m5.3: multiple regression D ~ A + M

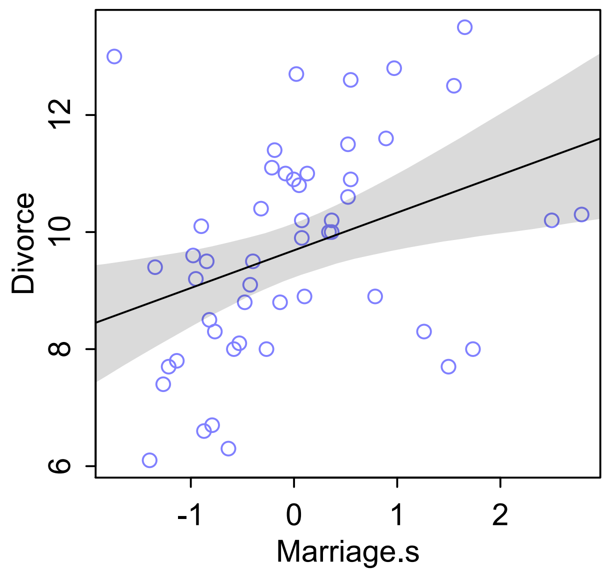
Multiple regression

	mean	sd	5.5%	94.5%
a	0.00	0.10	-0.16	0.16
bM	-0.07	0.15	-0.31	0.18
bA	-0.61	0.15	-0.85	-0.37
sigma	0.79	0.08	0.66	0.91

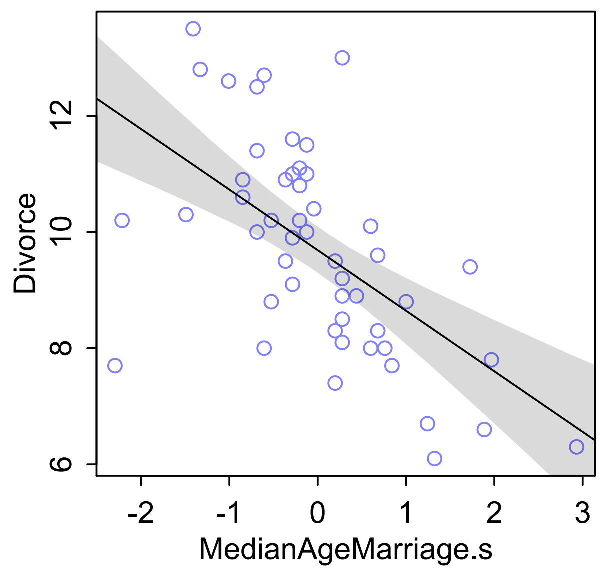


- Once we know median age marriage, little additional value in knowing marriage rate.
- Once we know marriage rate, still value in knowing median age marriage.
- If we *don't know* median age marriage, still useful to know marriage rate.

Posterior predictions



- Lots of plotting options now
  1. Predictor residual plots
  2. Counterfactual plots
  3. Posterior prediction plots



Masked association

- Sometimes association between outcome and predictor masked by another variable
- Need both variables to see influence of either
- Tends to arise when
  - Another predictor associated with outcome *in opposite direction*
  - Both predictors associated with one another
- Noise in predictors can also mask association (*residual confounding*)

Categorical variables

- Many predictors are discrete, unordered categories
  - Gender, region, species
- How to use in regression?
- Two approaches
  - Use *dummy/indicator* variables
  - Use index variables
- Most automated software uses dummy variables
- Usually easier to think & code with index variables

Index variable

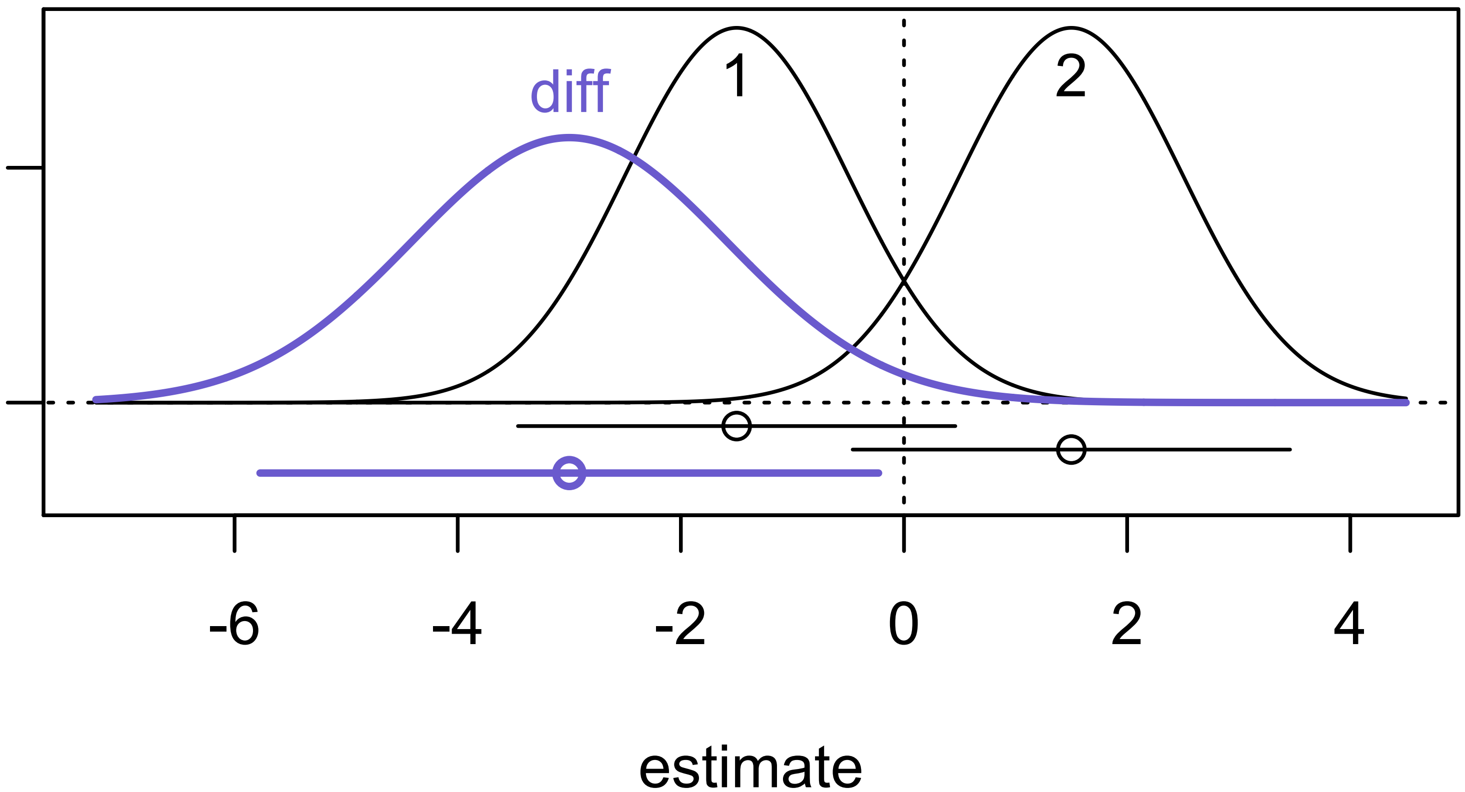
R code 5.36

```
d$sex <- ifelse( d$male==1 , 2 , 1 )
str( d$sex )
```

```
 num [1:544] 2 1 1 2 1 2 1 2 1 2 ...
```

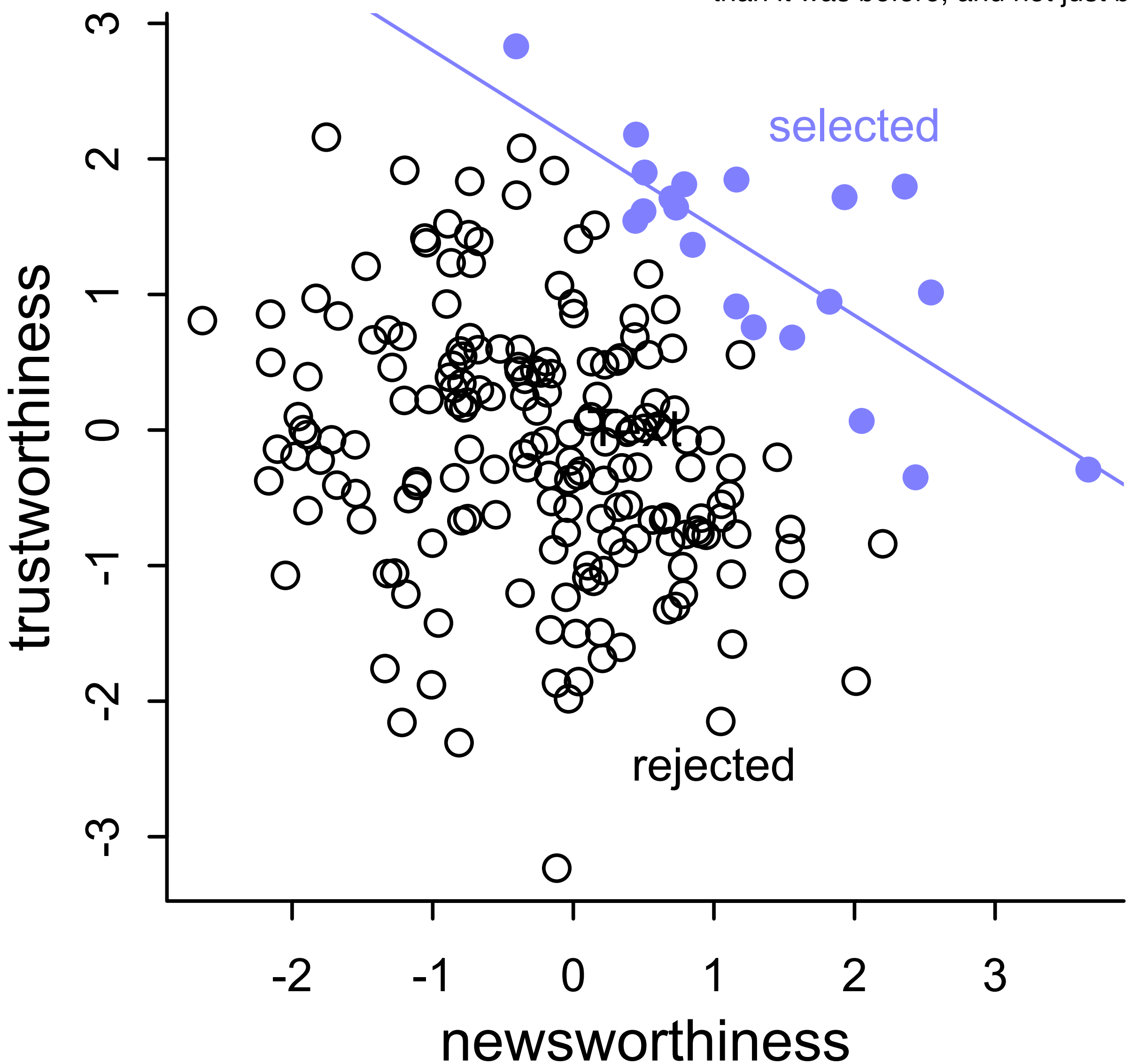
$$
\begin{aligned}
h_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\text{SEX}[i]} \\
\alpha_j &\sim \text{Normal}(178, 20) \quad \text{, for } j = 1..2 \\
\sigma &\sim \text{Uniform}(0, 50)
\end{aligned}
$$

Difference and uncertainty



Selection-distortion effect

when a sample is selected on a combination of 2 (or more) variables, the relationship between those 2 variables is different after selection than it was before, and not just because of restriction of range



Why not just add everything?

- Could just add all available predictors to model
  - "We controlled for…"
- Almost always a bad idea
  - Adding variables *creates* confounds
  - Residual confounding
  - Overfitting

The Fork

$$X \leftarrow Z \rightarrow Y$$

Open unless you condition on Z

The Pipe

$$X \rightarrow Z \rightarrow Y$$

Open unless you condition on Z

The Collider

$$X \rightarrow Z \leftarrow Y$$

Closed until you condition on Z

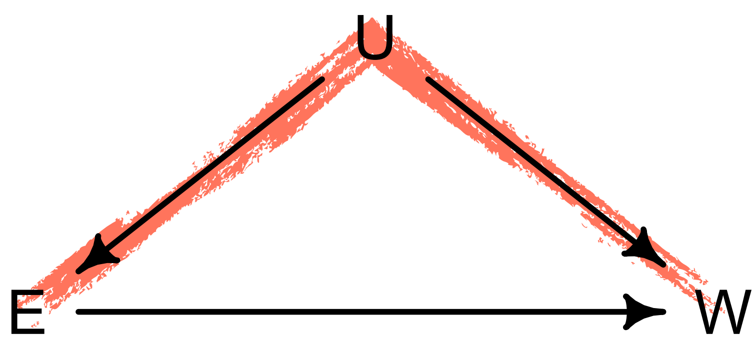
The Descendant

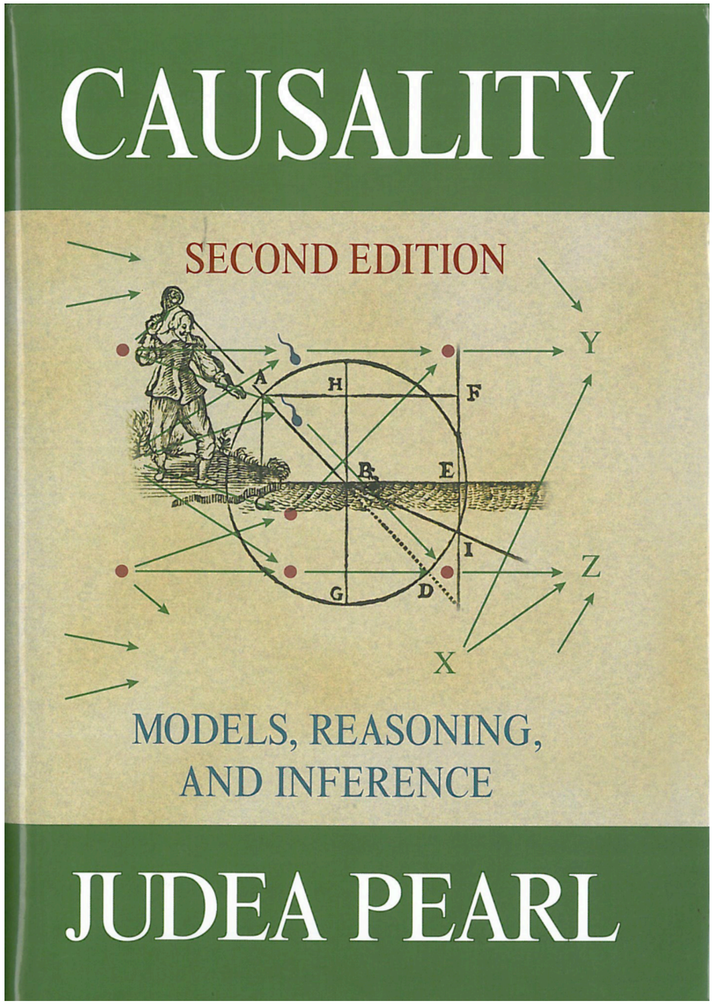
$$\begin{array}{ccccc} X & \rightarrow & Z & \rightarrow & Y \\ & & \downarrow & & \\ & & A & & \end{array}$$

Conditioning on A is like conditioning on Z

Shutting the back door

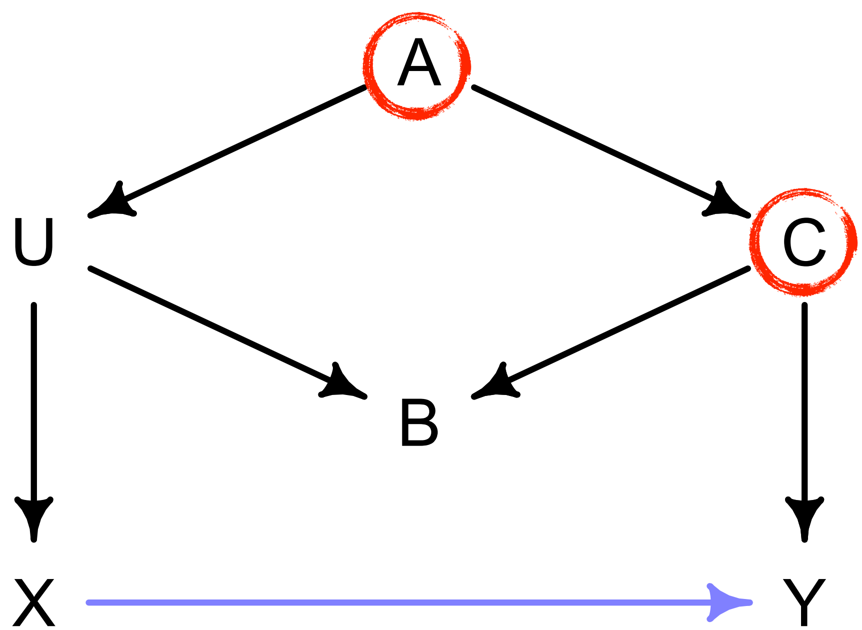
- What ties these examples together:
- The **back-door criterion**: Confounding caused by existence of open back door paths from X to Y
- If you know your elements, you know how to open/close each of them





Something more interesting

- Which variables, if any, should you condition on to infer X → Y?
- Condition on A or C. Do not condition on B.



(1) $X \leftarrow U \leftarrow A \rightarrow C \rightarrow Y$

This path is open.

(2) $X \leftarrow U \rightarrow B \leftarrow C \rightarrow Y$

This path is closed.

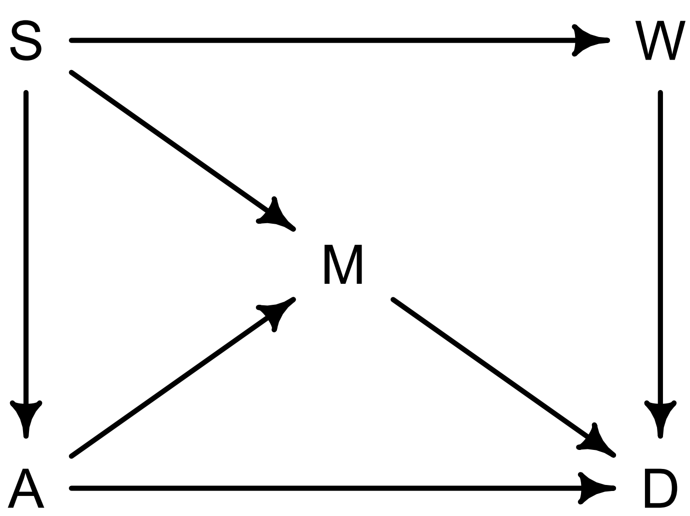
Implied conditional independence

- Given DAG, can test some implications

```
impliedConditionalIndependencies( dag_6.2 )
```

R code 6.36

```
A _||_ W | S
D _||_ S | A, M, W
M _||_ W | S
```



(1) A and W independent, conditioning on S
(2) D and S independent, conditioning on A, M, & W
(3) M and W independent, conditioning on S

Causal inference hard but possible

- Demonstrate capable of inferring cause
- Experiments not required!
- Experiments not always practical & ethical
  - Disease, evolution, development, dynamics of popular music, global climate, war
- Experiments must choose an intervention
  - Interventions influence many variables at once
  - Experimentally manipulate obesity?