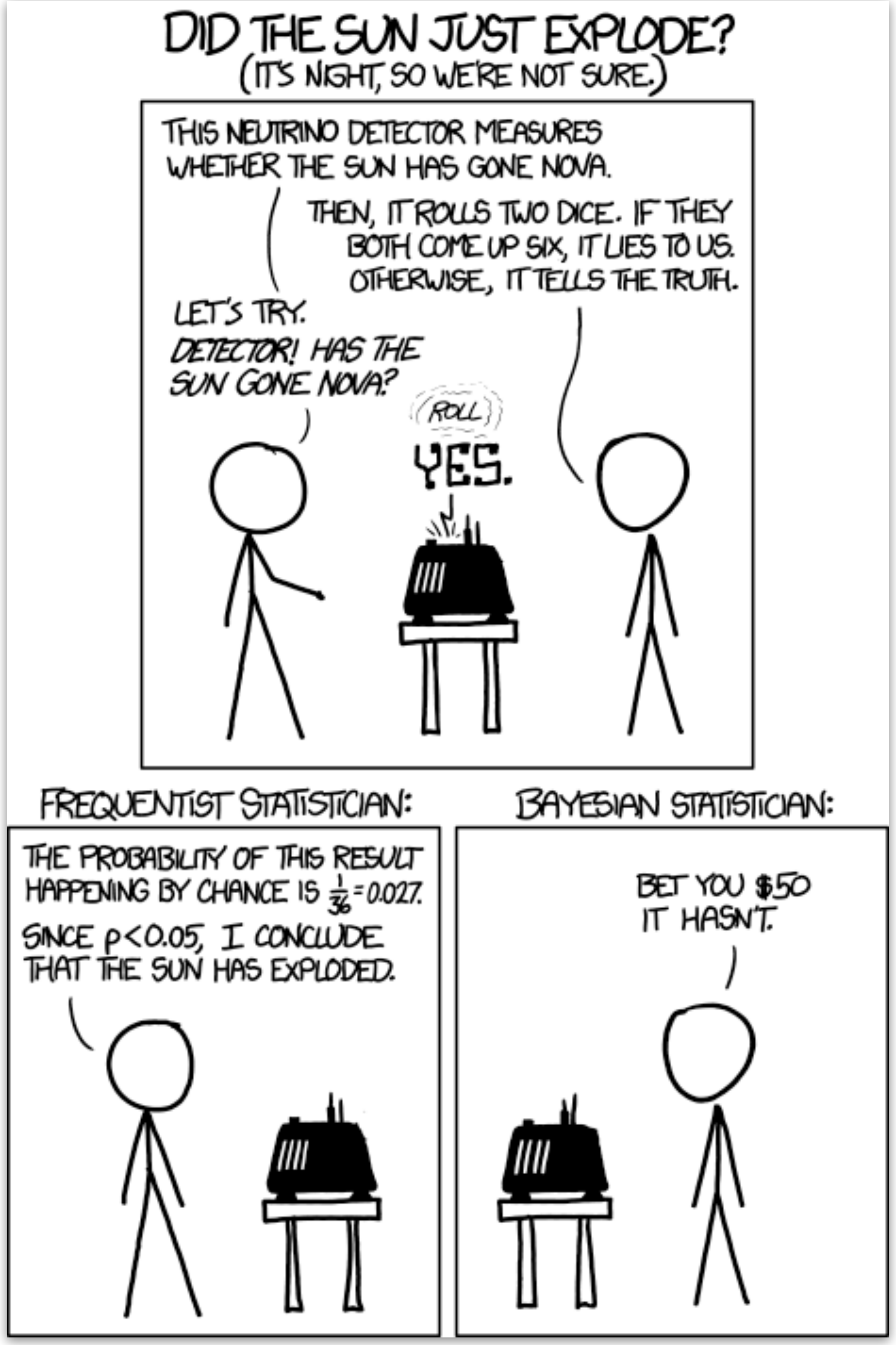
PSYCH 252: Statistical Methods for Behavioral and Social Sciences

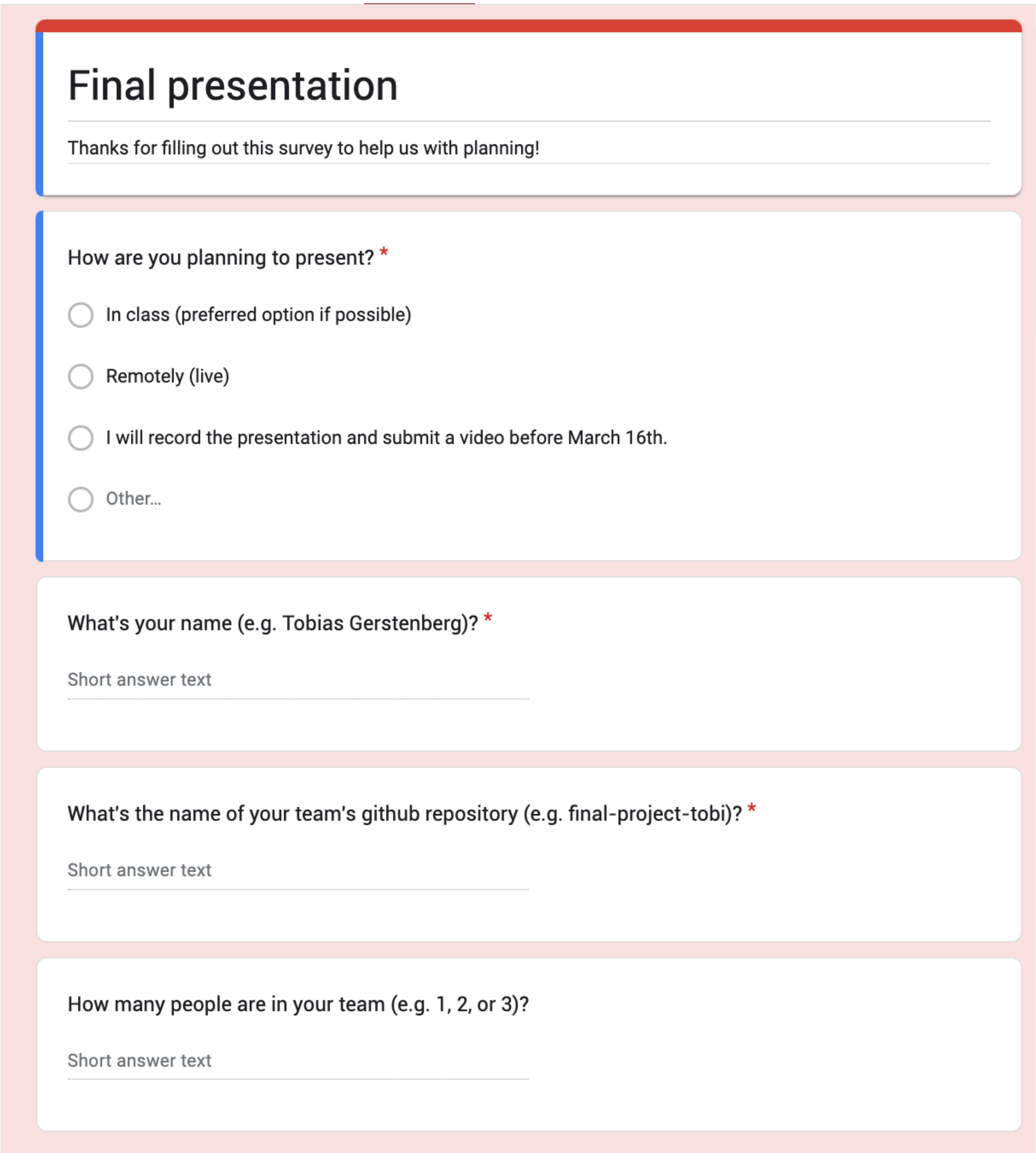
Bayesian data analysis 1



03/03/2025

Logistics

Final presentations



Final presentation

Thanks for filling out this survey to help us with planning!

How are you planning to present? *

- In class (preferred option if possible)
- Remotely (live)
- I will record the presentation and submit a video before March 16th.
- Other...

What's your name (e.g. Tobias Gerstenberg)? *

Short answer text

What's the name of your team's github repository (e.g. final-project-tobi)? *

Short answer text

How many people are in your team (e.g. 1, 2, or 3)?

Short answer text

https://tinyurl.com/psych252presentation25

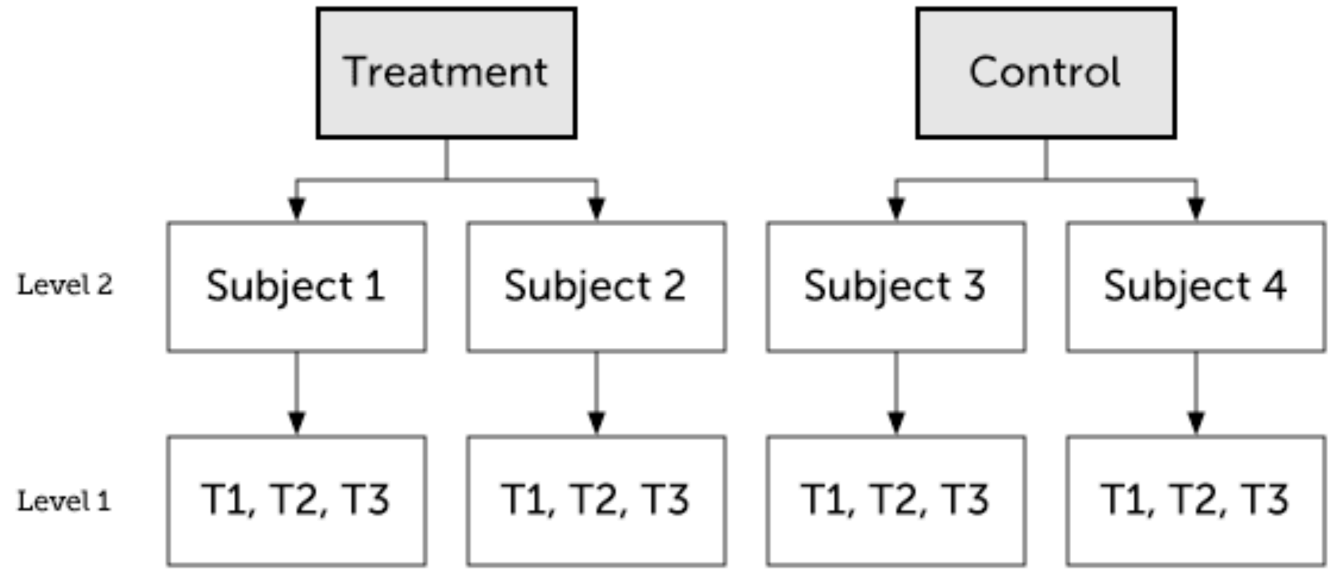
Plan for today

- Quick recap
- Causality
  - Patterns of inference
  - Should I control?
  - Mediation
  - Moderation
- Bayesian data analysis
  - Comparison between frequentist and Bayesian data analysis
  - Quick flash from the past
  - Flipping coins
  - What affects the posterior?
  - Ingredients: likelihood, prior, inference
  - Doing Bayesian data analysis

Quick recap

Quick recap

Graphical representation



assume a linear effect of time

Simulate data

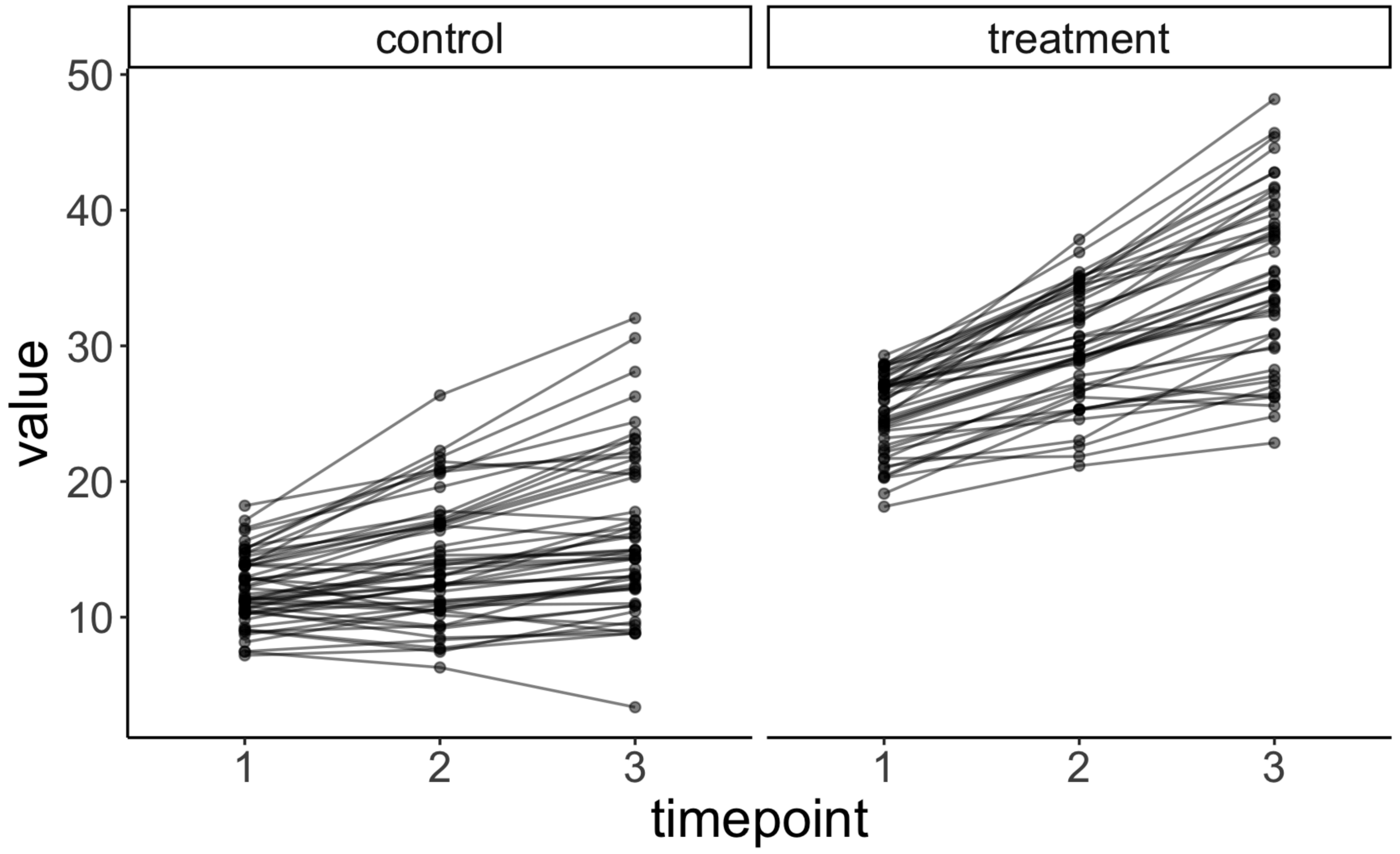
	participant	timepoint	intercept_participant	time_participant	condition	value
1	1	1	-1.252907621	-0.310183339	1	20.573131
2	1	2	-1.252907621	-0.310183339	1	22.844077
3	1	3	-1.252907621	-0.310183339	1	24.367254
4	2	1	0.367286648	0.021057937	0	12.119622
5	2	2	0.367286648	0.021057937	0	15.062837
6	2	3	0.367286648	0.021057937	0	17.513515
7	3	1	-1.671257225	-0.455460824	1	17.470186
8	3	2	-1.671257225	-0.455460824	1	22.446022
9	3	3	-1.671257225	-0.455460824	1	24.248006
10	4	1	3.190561604	0.079014386	1	24.844308
11	4	2	3.190561604	0.079014386	1	28.220589
12	4	3	3.190561604	0.079014386	1	28.880339
13	5	1	0.659015544	-0.327292322	1	22.047393
14	5	2	0.659015544	-0.327292322	1	25.189133
15	5	3	0.659015544	-0.327292322	1	28.051351

```
set.seed(1)

n_participants = 100
n_timepoints = 3
n_conditions = 2
p_condition = 0.5
b0 = 10 # intercept
b1 = 10 # condition
b2 = 2 # time
b3 = 3 # interaction
sd_intercept_participant = 2
sd_time_participant = 2
sd_residual = 1

df.data = tibble(participant = rep(1:n_participants, each = n_timepoints),
                 timepoint = rep(1:n_timepoints, times = n_participants),
                 intercept_participant = rep(rnorm(n_participants, sd = sd_intercept_participant),
                                             each = n_timepoints),
                 time_participant = rep(rnorm(n_participants, sd = sd_time_participant),
                                        each = n_timepoints)) %>%
  group_by(participant) %>%
  mutate(condition = rbinom(n = 1, size = 1, prob = p_condition)) %>%
  ungroup() %>%
  mutate(value = b0 + intercept_participant +
           b1 * condition +
           (b2  + time_participant) * timepoint +
           b3 * condition * timepoint +
           rnorm(n_participants * n_timepoints, sd = sd_residual))
```

random slopes

21

Plot data



22

Fit the model

```
fit = lmer(formula = value ~ 1 + condition * timepoint + (1 + timepoint | participant),
           data = df.data)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: value ~ 1 + condition * timepoint + (1 + timepoint | participant)
   Data: df.data

REML criterion at convergence: 1360.3

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.14633 -0.46360  0.03902  0.42302  2.82945

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 participant (Intercept) 3.190    1.786
             timepoint   3.831    1.957    -0.06
 Residual                1.149    1.072
Number of obs: 300, groups:  participant, 100

Fixed effects:
                    Estimate Std. Error t value
(Intercept)          10.0101     0.3328  30.079
condition            10.0684     0.4854  20.741
timepoint             2.0595     0.2883   7.143
condition:timepoint   2.9090     0.4205   6.917

Correlation of Fixed Effects:
            (Intr) condtn timpnt
condition   -0.686
timepoint   -0.266  0.182
cndtn:tmpnt  0.182 -0.266 -0.686
```

23

Quick recap

What shall I include as random effects?

- mixed opinions on the topic
- go maximal!

Journal of Memory and Language 68 (2013) 255–278

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Random effects structure for confirmatory hypothesis testing: Keep it maximal

Dale J. Barr[a,*], Roger Levy[b], Christoph Scheepers[a], Harry J. Tily[c]

[a] Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead St., Glasgow G12 8QB, United Kingdom
[b] Department of Linguistics, University of California at San Diego, La Jolla, CA 92093-0108, USA
[c] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

"Through theoretical arguments and Monte Carlo simulation, we show that LMEMs generalize best when they include the maximal random effects structure justified by the design. ...

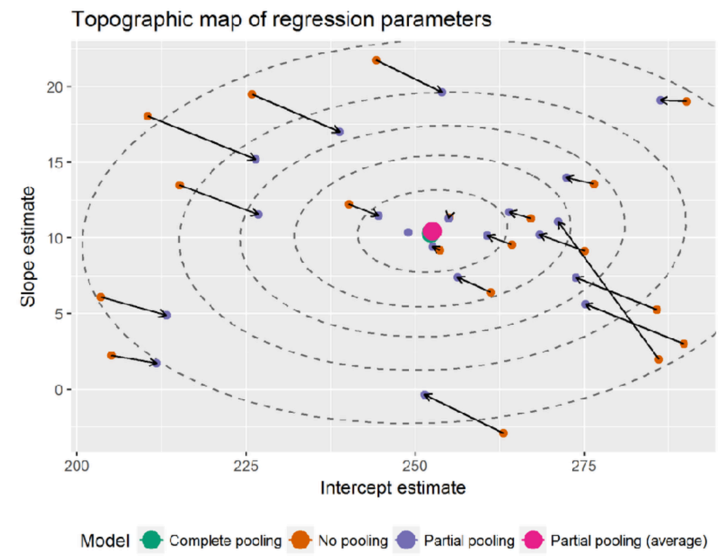
Maximal LMEMs should be the 'gold standard' for confirmatory hypothesis testing in psycholinguistics and beyond."

Remove the correlation component from your model

```r
# fit the model
fit.lmer = lmer(formula = reaction ~ 1 + days + (1 + days | subject),
                data = df.sleep)
# model summary
fit.lmer %>%
  summary()
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: reaction ~ 1 + days + (1 + days | subject)
   Data: df.sleep

REML criterion at convergence: 1771.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.9707 -0.4703  0.0276  0.4594  5.2009

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 subject  (Intercept) 582.73   24.140
          days         35.03    5.919   0.07
 Residual             649.36   25.483
Number of obs: 183, groups:  subject, 20

Fixed effects:
            Estimate Std. Error t value
(Intercept)  252.543      6.433  39.256
days          10.452      1.542   6.778

Correlation of Fixed Effects:
     (Intr)
days -0.137
```

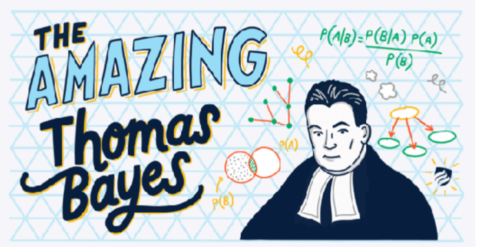
multivariate Gaussian



33

What if lmer() fails to converge?

1. We drop random effects in the following order: random correlations, random slopes of covariates (where significance is of no interest), random intercepts ("0+" instead "1+") (following Barr et al., 2013). We never remove the random slopes of the variables of interest (i.e., the ones for which we want to conduct significance tests).
   Please note that removing random correlation terms can be tricky if random slopes are estimated for factors with 3 or more levels. In that case, it is probably easiest to use `afex::mixed()` with `expand_re = TRUE` (an alternative option is to create manually the relevant contrasts yourself and add them as predictors to your model, which allows you to suppress the random corrections using the double pipe symbol ||).
2. We try to run separate analyses: For example, one model to only test the fixed and random effect of A (with fixed effect of B present); then one model to only test the effect of B.
   If we really have to drop random slopes, we follow the next step:
3. We follow the PCA approach suggested by **rePsychLing** (see Bates et al., 2015) that is performing a PCA on the random effects and following the guidelines described in the paper.
   a. We use a likelihood ratio test to test whether the model fit becomes significantly worse. As we prefer a more conservative approach here (i.e., rather err on the side of keeping too many random effects; we prioritize avoiding inflated Type 2 errors for this kind of decision), we use larger alpha-level of .2 (Matuschek et al., 2017).
   b. Alternatively, we suggest an Information criterion approach to avoid using a *p* value for our inclusion/exclusion decision, but choose the best model based on *BIC* or *AIC*.



3.2.2. Or we choose a Bayesian approach

As an alternative to targeting convergence issues within **lme4**, we suggest fitting the same model with **brms** and comparing it to the **lme4** fit. We assume that both provide similar results when

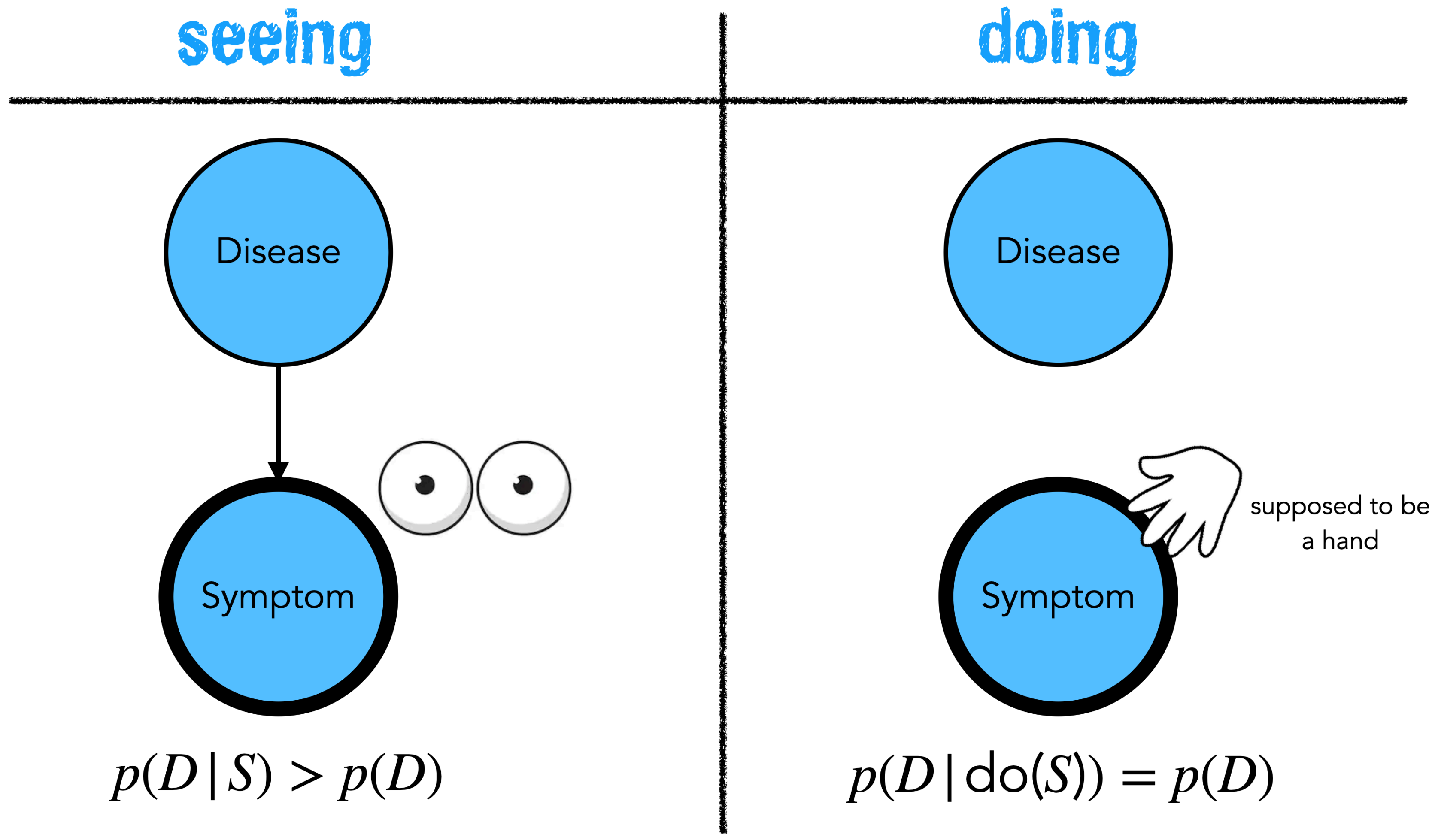
37

Quick recap

correlation does not imply causation

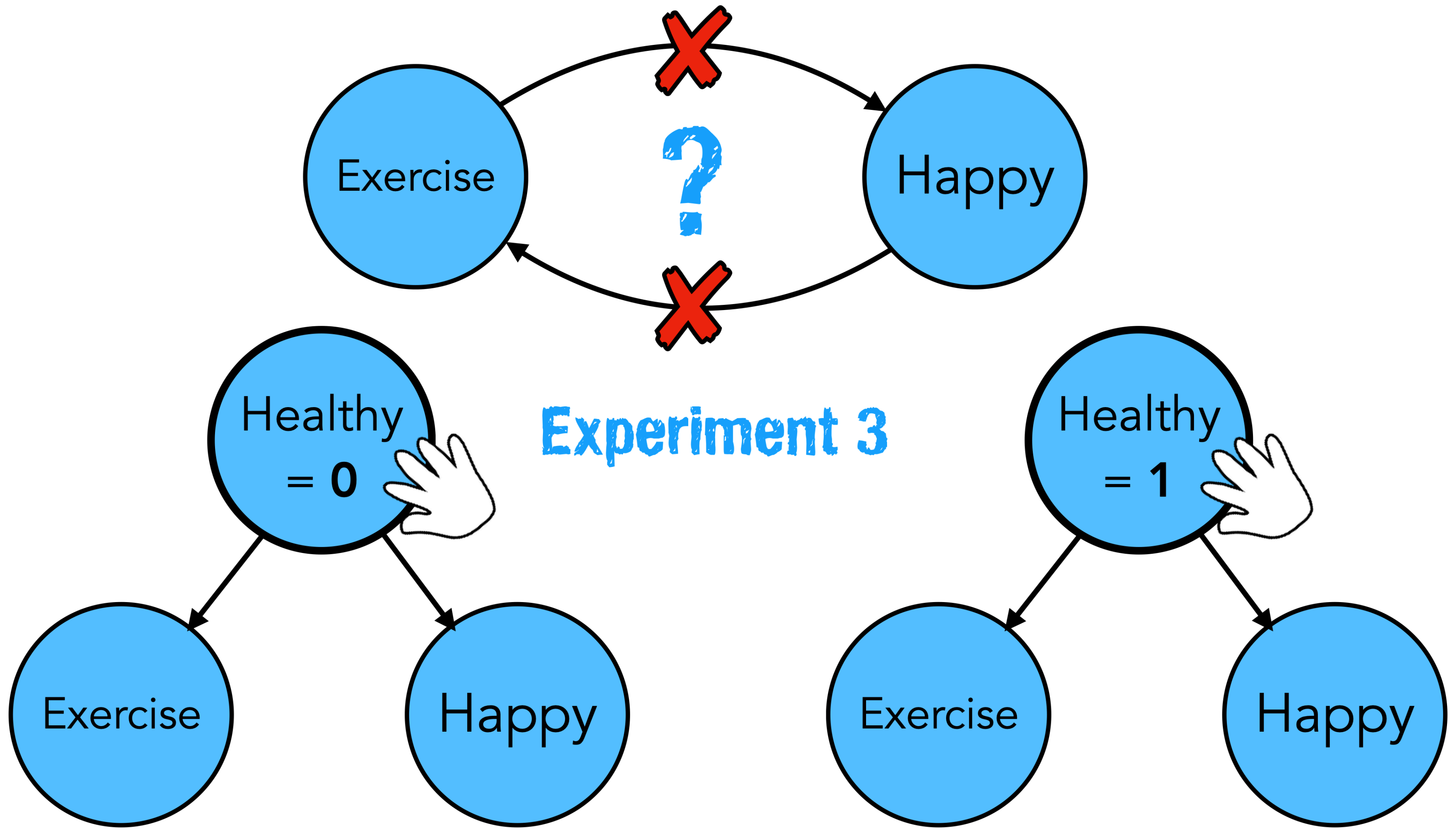
40

Observation vs. Intervention



$p(D \mid S) > p(D)$ $\quad$ $p(D \mid \text{do}(S)) = p(D)$

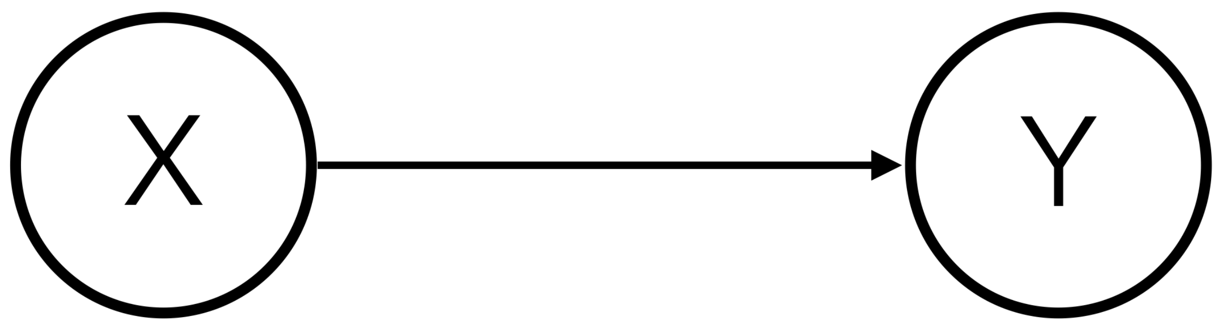
Inferring causal structure through intervention



$p(\text{Happy} \mid \text{do}(\text{Healthy} = 0)) = 0.05$ $\quad$ $p(\text{Happy} \mid \text{do}(\text{Healthy} = 1)) = 0.5$

$p(\text{Exercise} \mid \text{do}(\text{Healthy} = 0)) = 0.1$ $\quad$ $p(\text{Exercise} \mid \text{do}(\text{Healthy} = 1)) = 0.75$

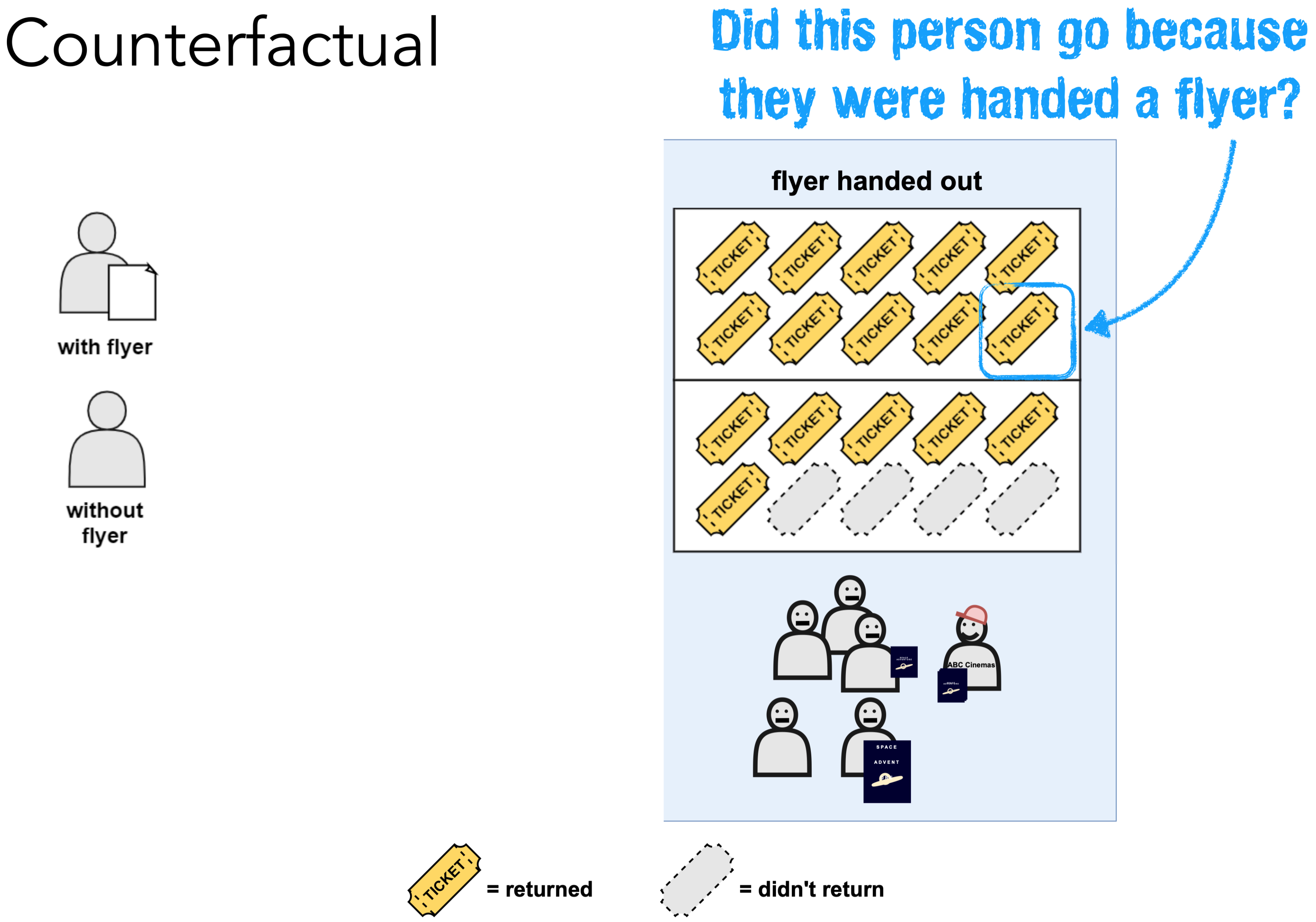
Quick recap



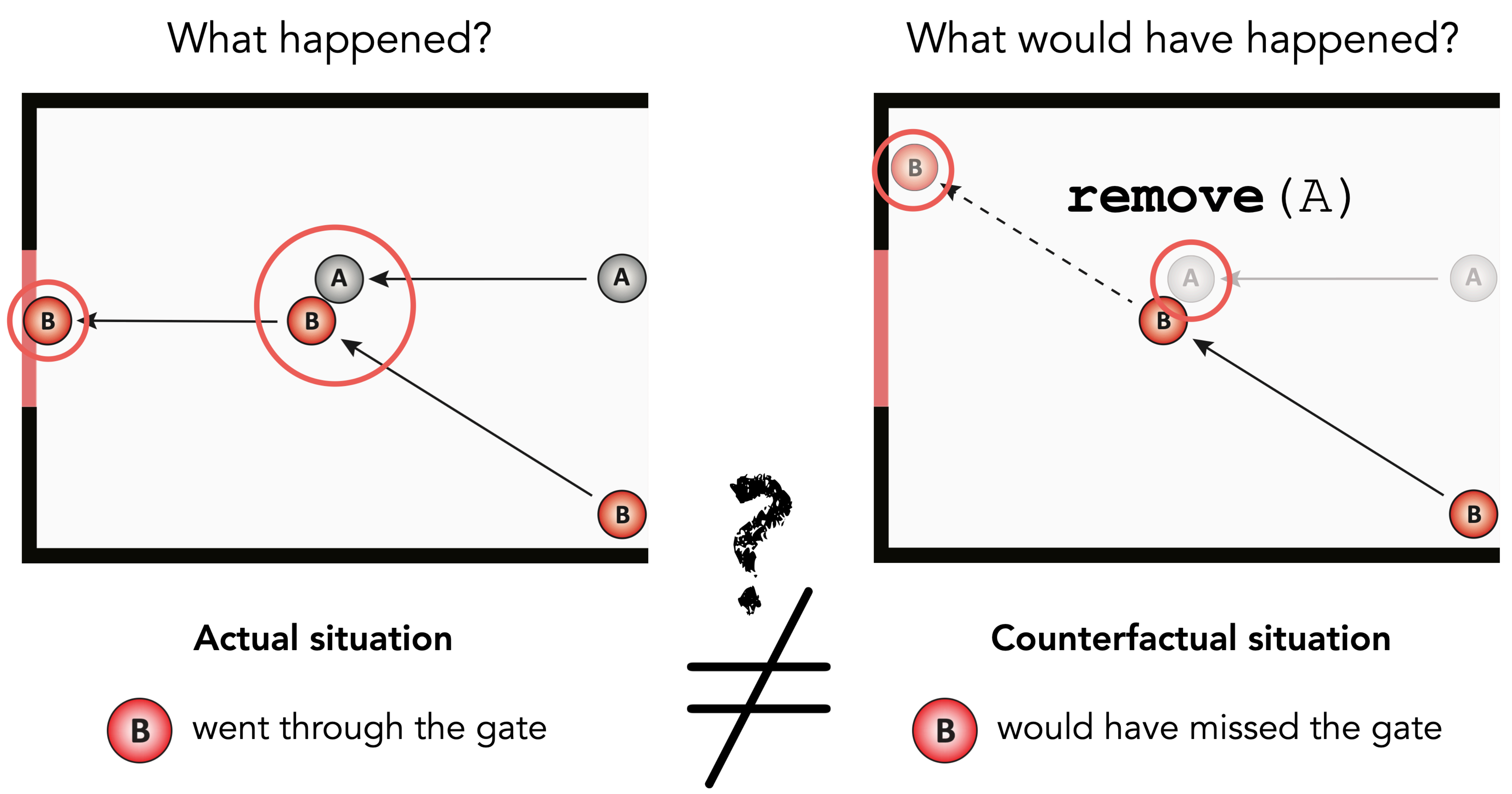
Level	Concept	Expression	Activity	Question	Example
I	Observation/ Prediction	$p(y\|x)$	Seeing	How does x change my belief in y?	Would the grass be wet if we *found* the sprinkler off?
II	Intervention/ Hypothetical	$p(y\|\mathrm{do}(x))$	Doing	Would y happen if I did x?	Would the the grass be wet if *made sure* that the sprinkler was off?
III	Counterfactual	$p(y_x\|x', y')$	Explaining	Would y have happened instead of y', if I had done x instead of x'?	Would the grass have been wet if we *had made sure* that the sprinkler was off, given that the grass is wet and the sprinkler on?

54

Counterfactual



Counterfactual Simulation Model



Patterns of inference

Patterns of inference: **Common cause**

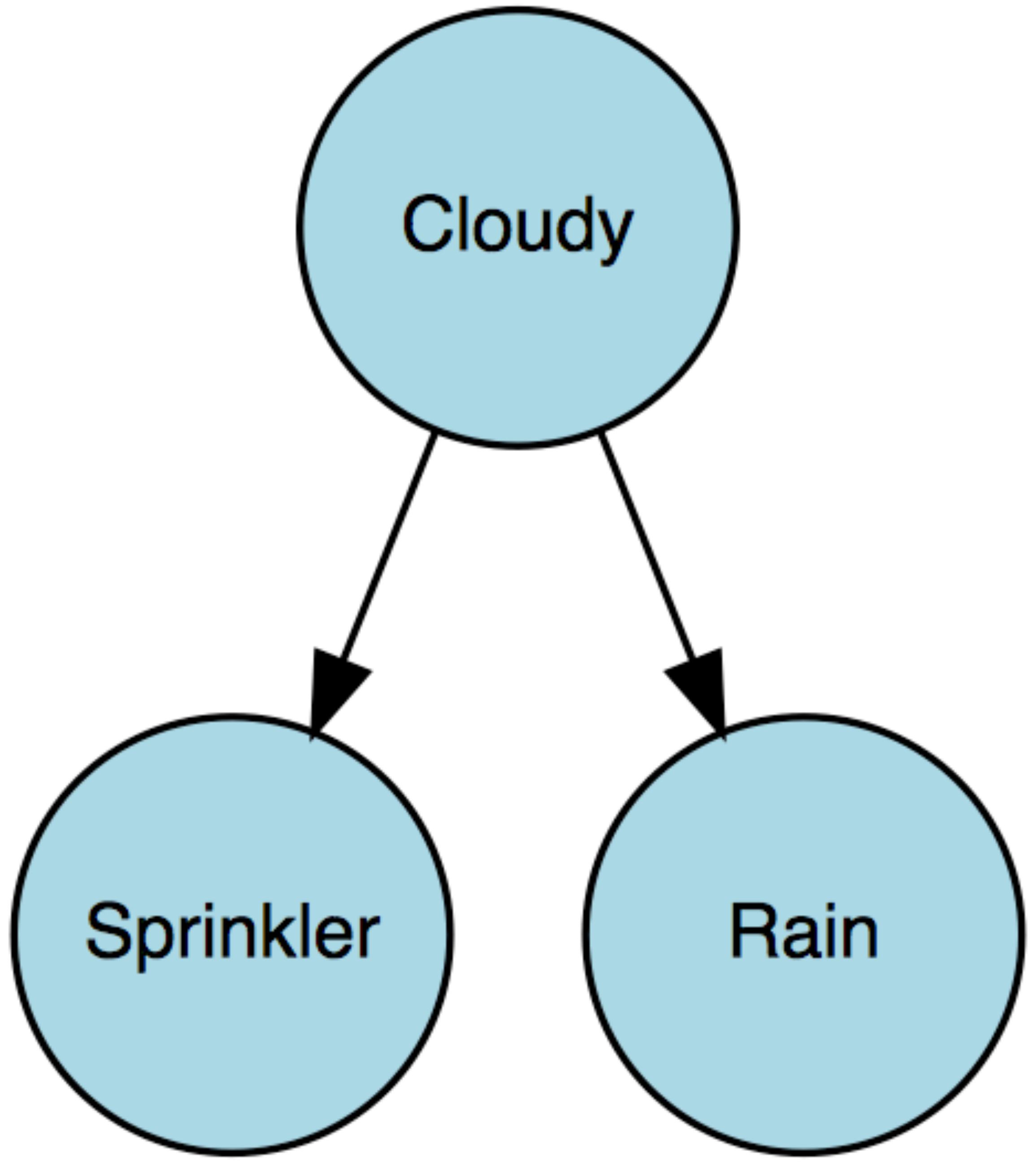
$$p(S \mid R) = p(S)$$

or

$$p(S \mid R) \neq p(S)$$

?

p(C)
0.5



C	p(S)	C	p(R)
F	0.5	F	0
T	0.1	T	0.3

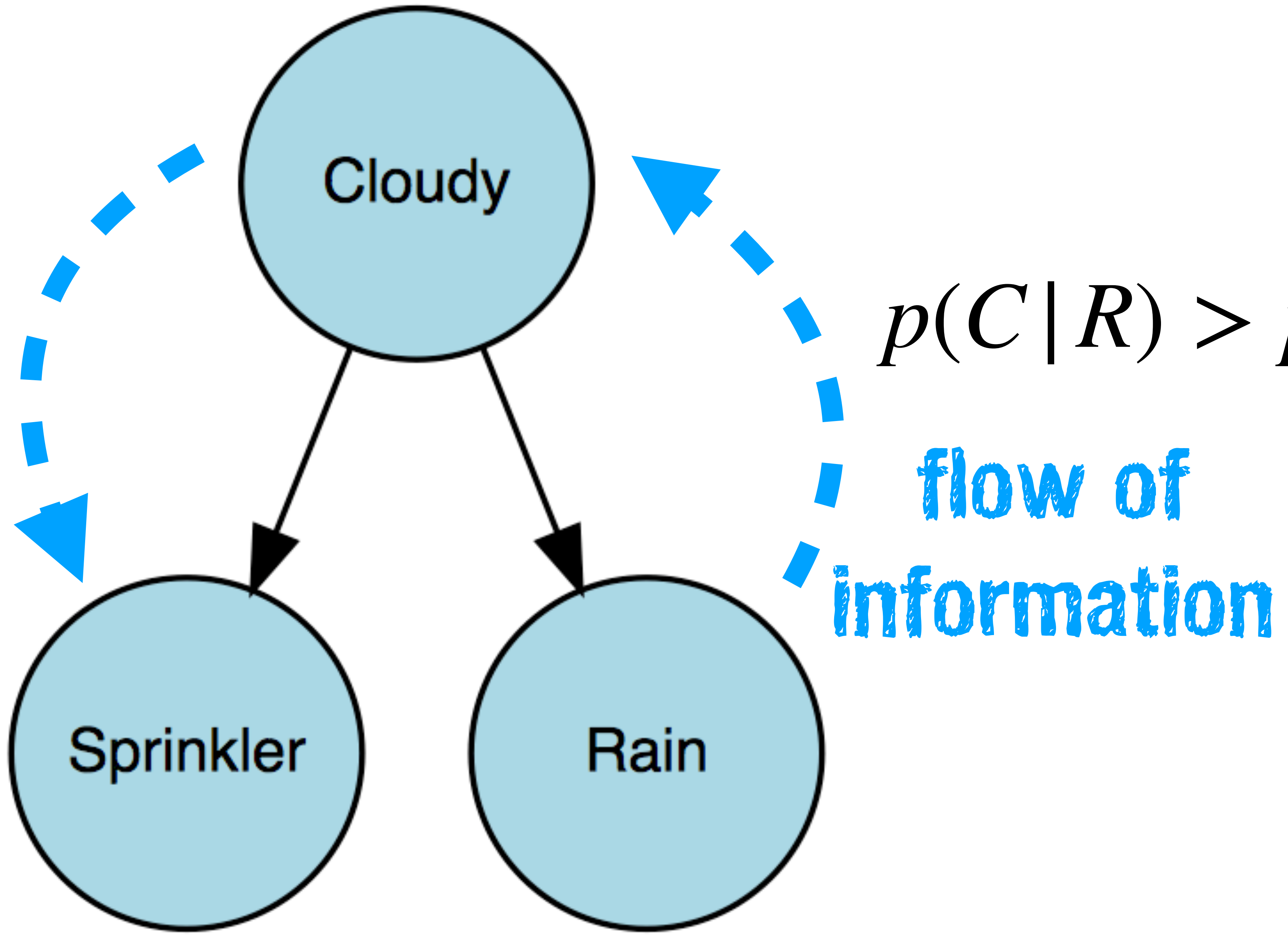
Patterns of inference: **Common cause**

- effects of a common cause are *unconditionally dependent*

$$p(S \mid R) \neq p(S)$$

p(C)
0.5

$$p(S \mid C) < p(S)$$



$$p(C \mid R) > p(C)$$

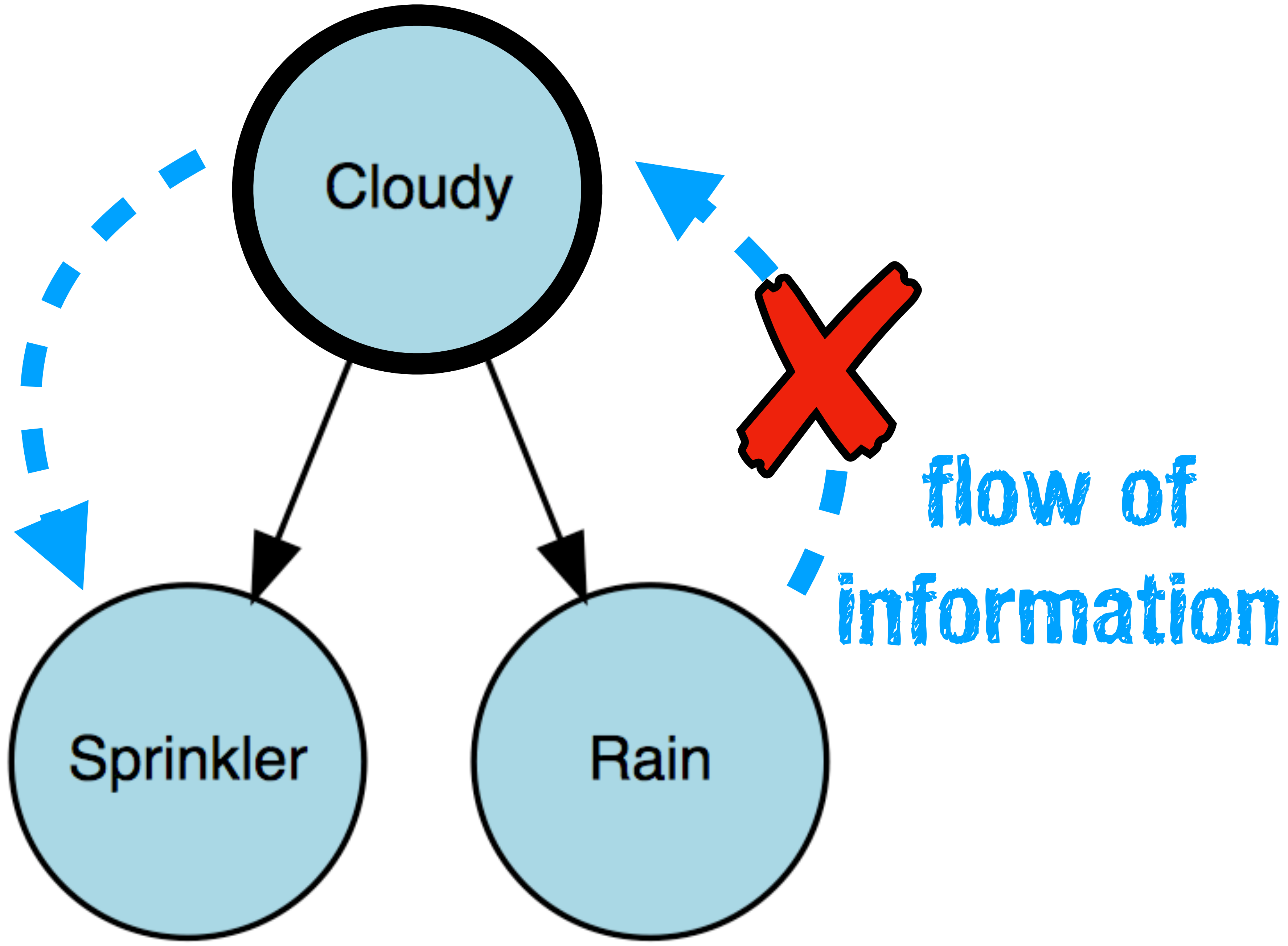
C	p(S)	C	p(R)
F	0.5	F	0
T	0.1	T	0.3

Patterns of inference: **Common cause**

- effects of a common cause are *conditionally independent given the cause*

$$p(S \mid R, C) = p(S \mid C)$$

p(C)
1

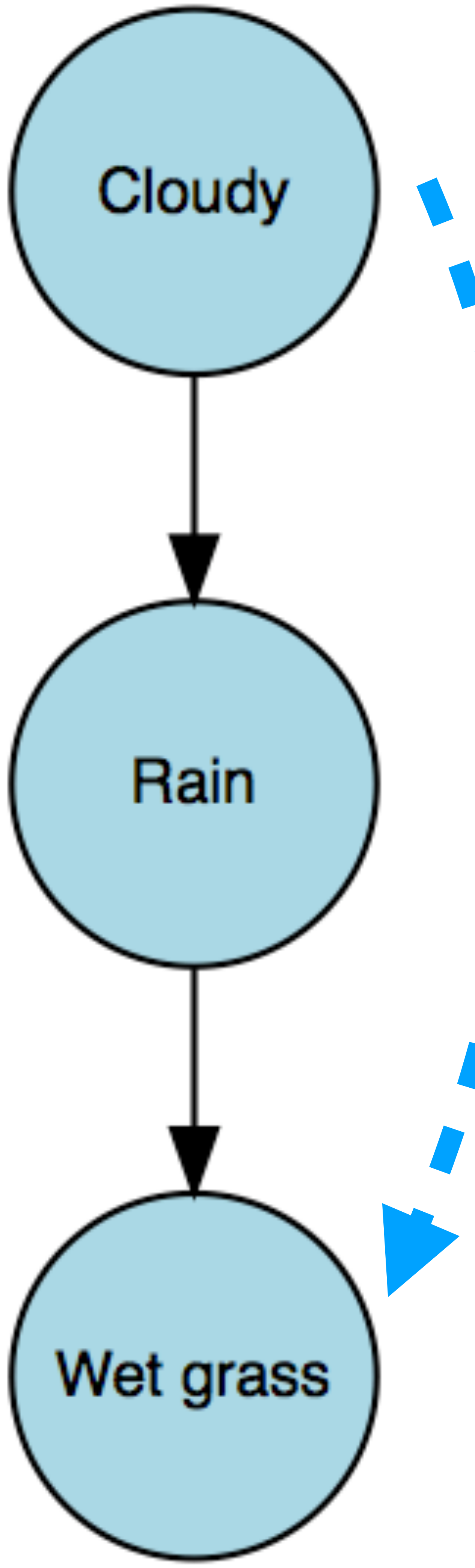


C	p(S)	C	p(R)
F	0.5	F	0
T	0.1	T	0.3

Patterns of inference: **Causal chain**

- cause and effect in a causal chain are *unconditionally dependent*

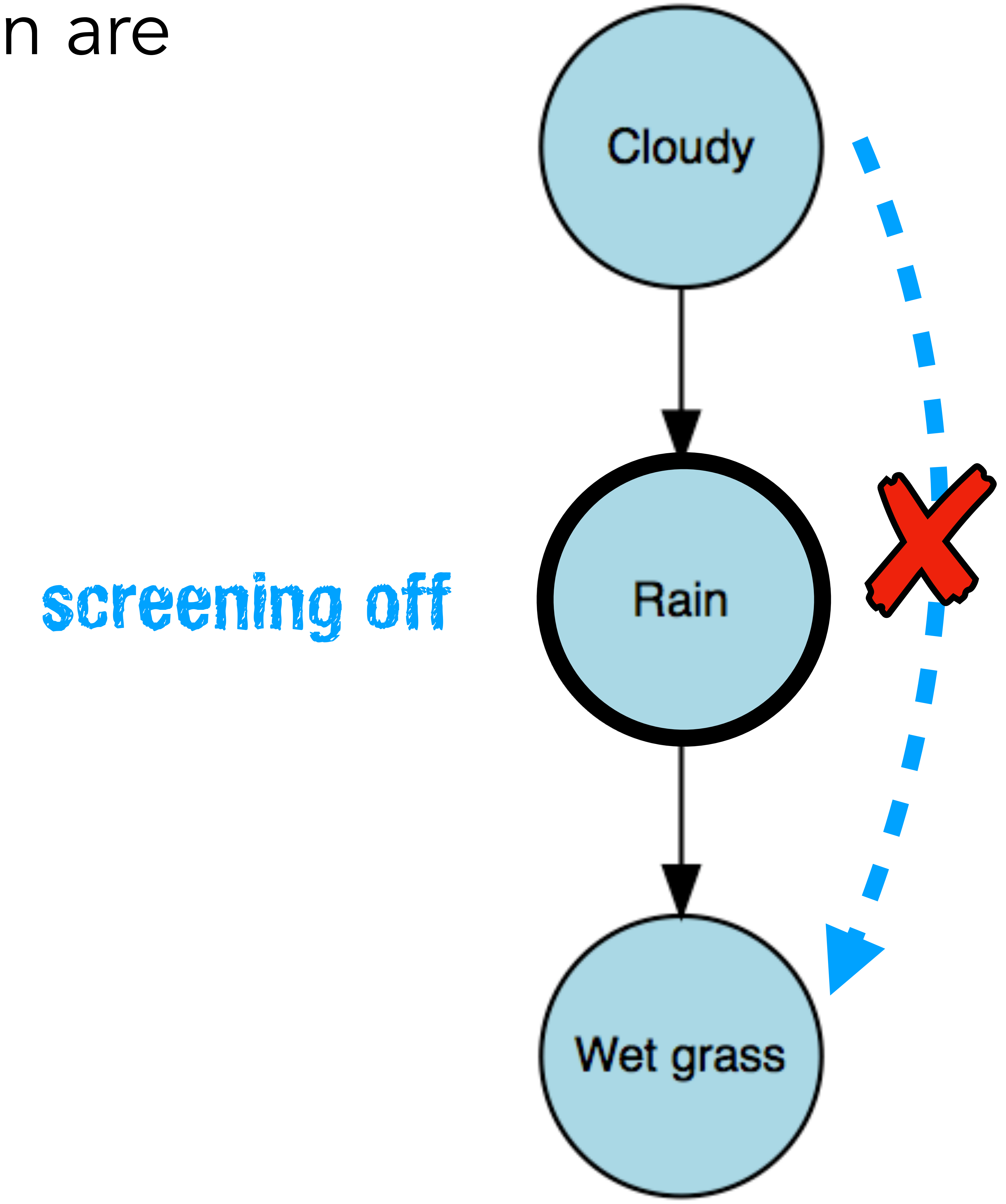
$$p(W \mid C) \neq p(W)$$



Patterns of inference: **Causal chain**

- cause and effect in a causal chain are *conditionally independent*

$$p(W \mid C, R) = p(W \mid R)$$

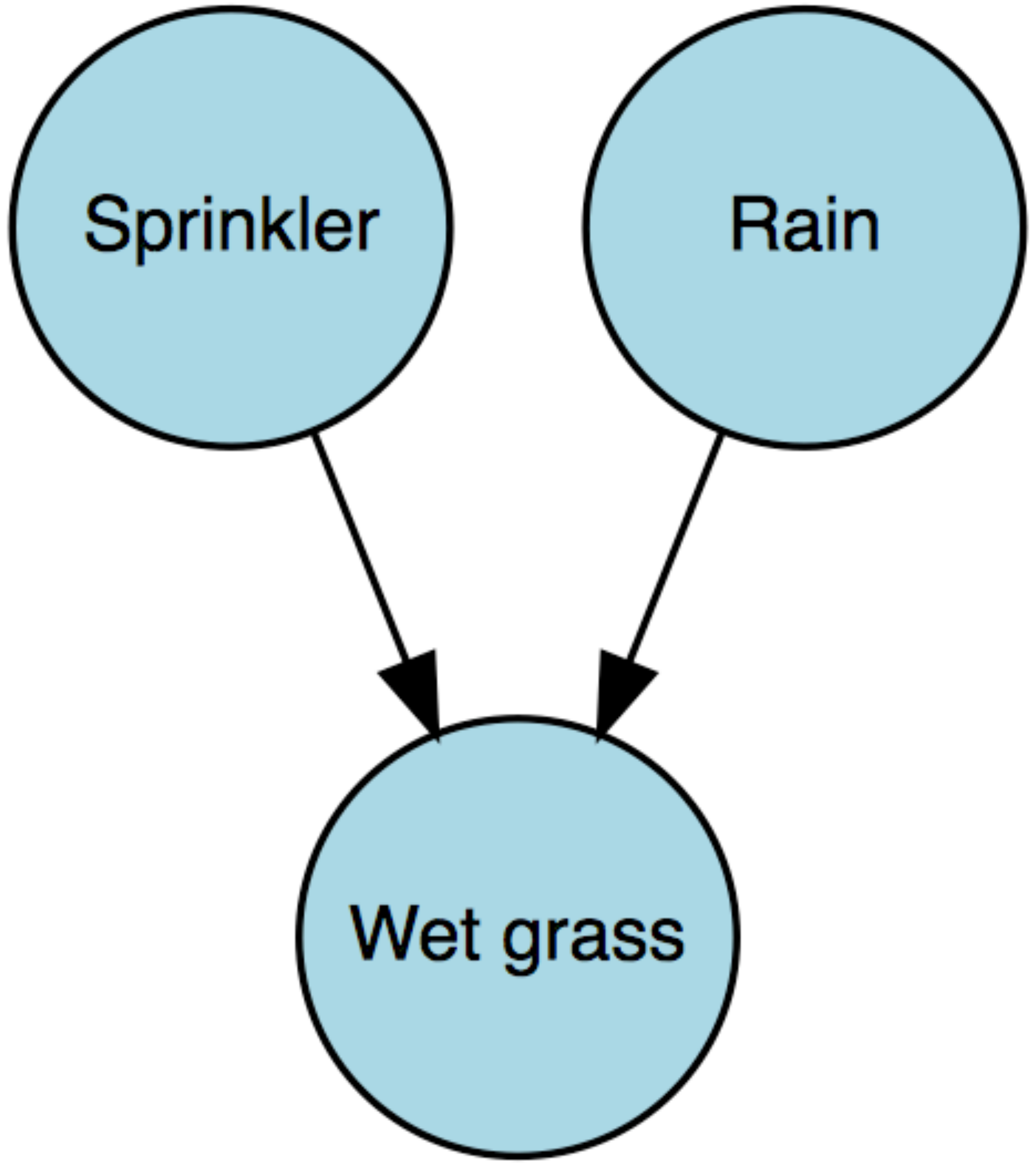


Patterns of inference: **Common effect**

- two causes of a common effect are *unconditionally independent*

$$p(S|R) = p(S)$$

(e.g. Sprinkler is set by a timer)



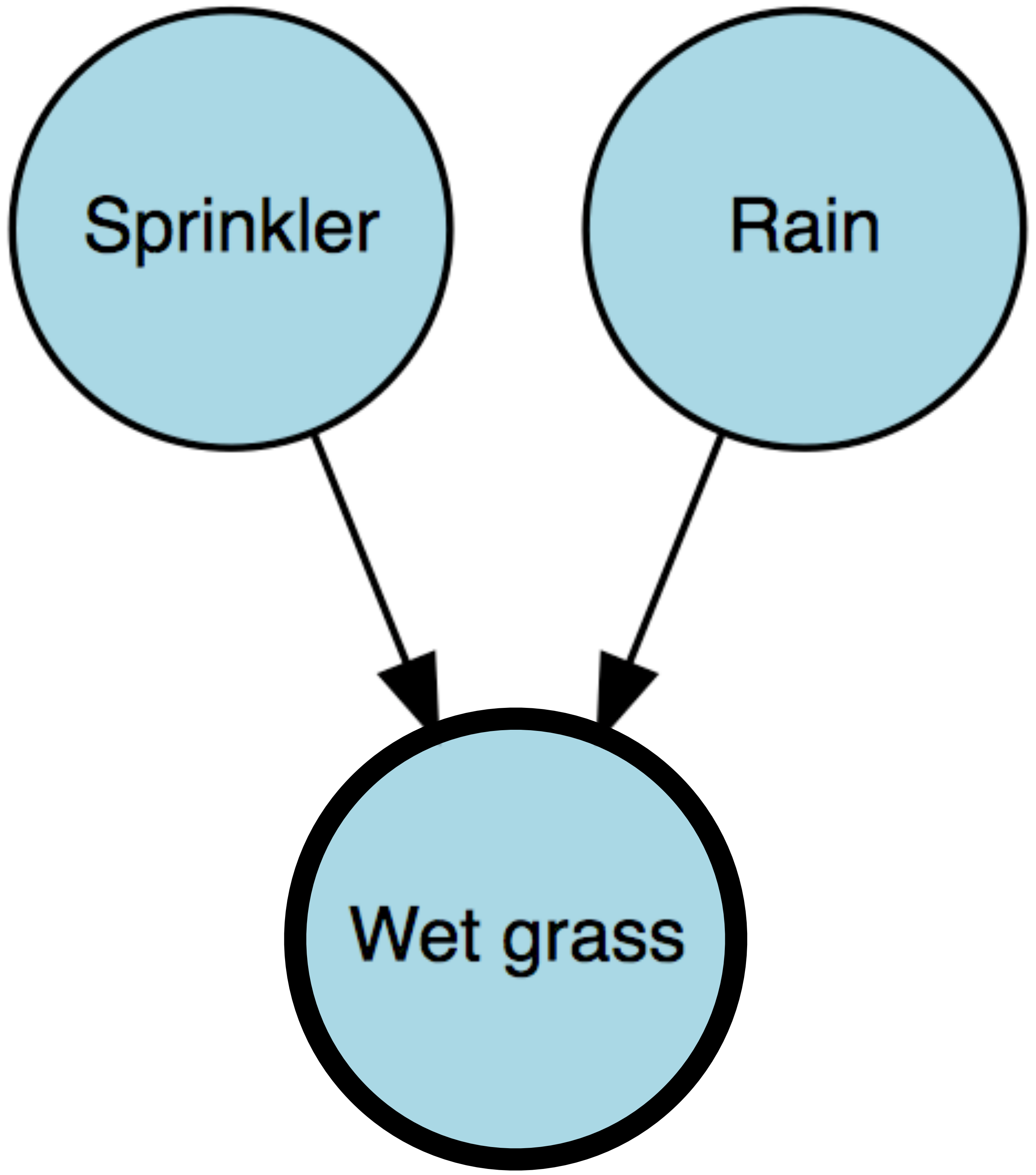
Patterns of inference: **Common effect**

- two causes of a common effect are *conditionally dependent given the effect*

$$p(S \mid R, W) \neq p(S \mid W)$$

- intuitively: both causes compete to explain the effect

explaining away



Note: The pattern of inference depends on the structural form which captures how `Sprinkler` and `Rain` jointly affect `Wet grass`. Explaining away holds for the commonly used noisy-or integration function.

Should I control?

When should I control for variables?

recent advances in graphical models have produced a way to help distinguish good from bad controls

***d*-separation**

directional

decide from a causal graph whether a set of variables `X` is independent of another set `Y`, given a third set `Z`

Goal: we want a precise (and unbiased) estimate of the predictive relationship between `X` and `Y`

we want to block all other paths from X to Y

http://bayes.cs.ucla.edu/BOOK-09/ch11-1-2-final.pdf

When should I control for variables?

How can I tell whether two variables are independent?

Recipe for independence

1. Draw the ancestral graph
2. "Moralize" the graph by "marrying" the parents
3. "Disorient" the graph by replacing arrows with edges
4. Delete the givens and their edges
5. Read the answer off the graph

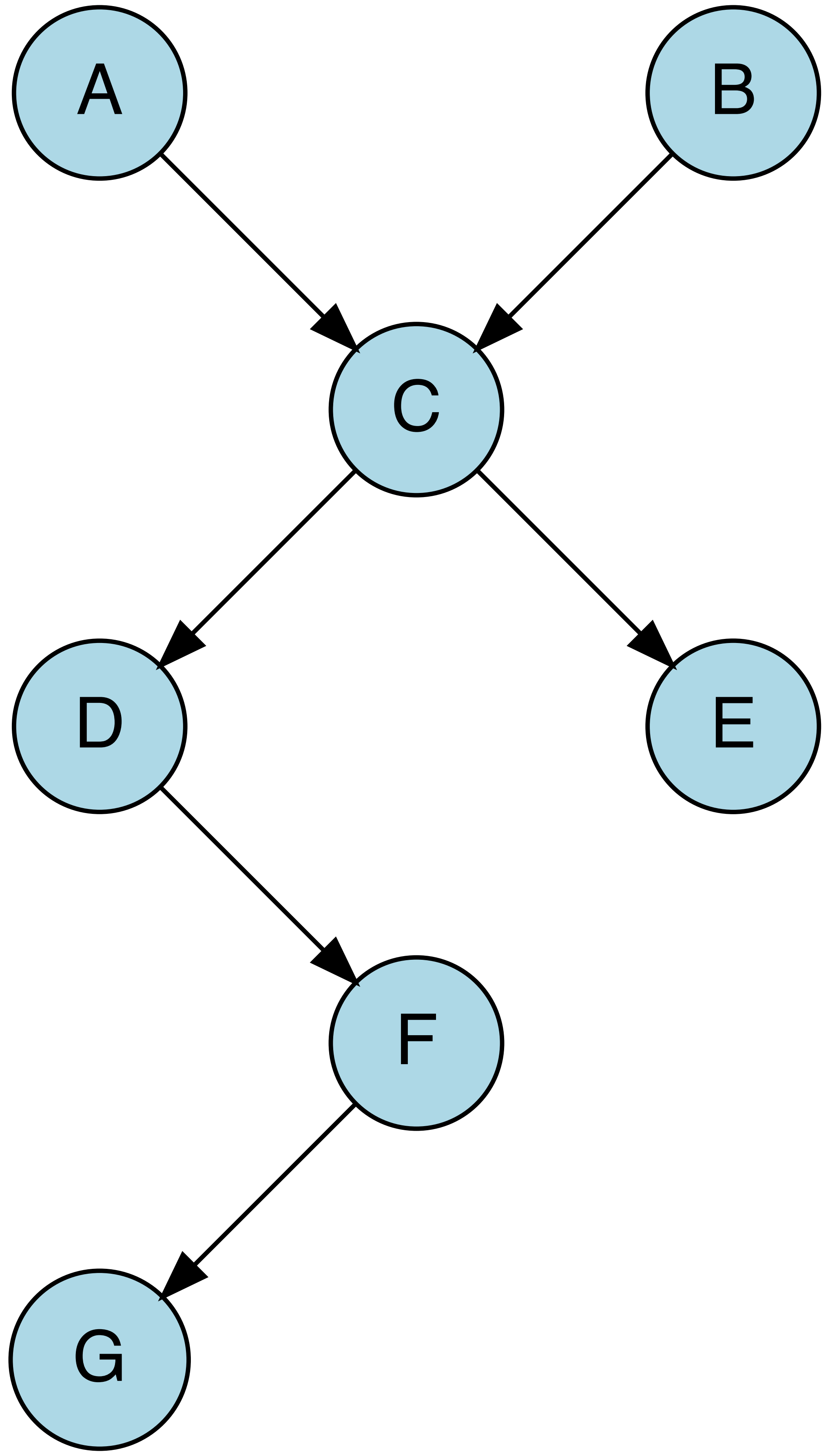
- if variables are **disconnected** they are independent
- if variables are connected (have a path between them) they are not guaranteed to be independent

http://web.mit.edu/jmn/www/6.034/d-separation.pdf

When should I control for variables?

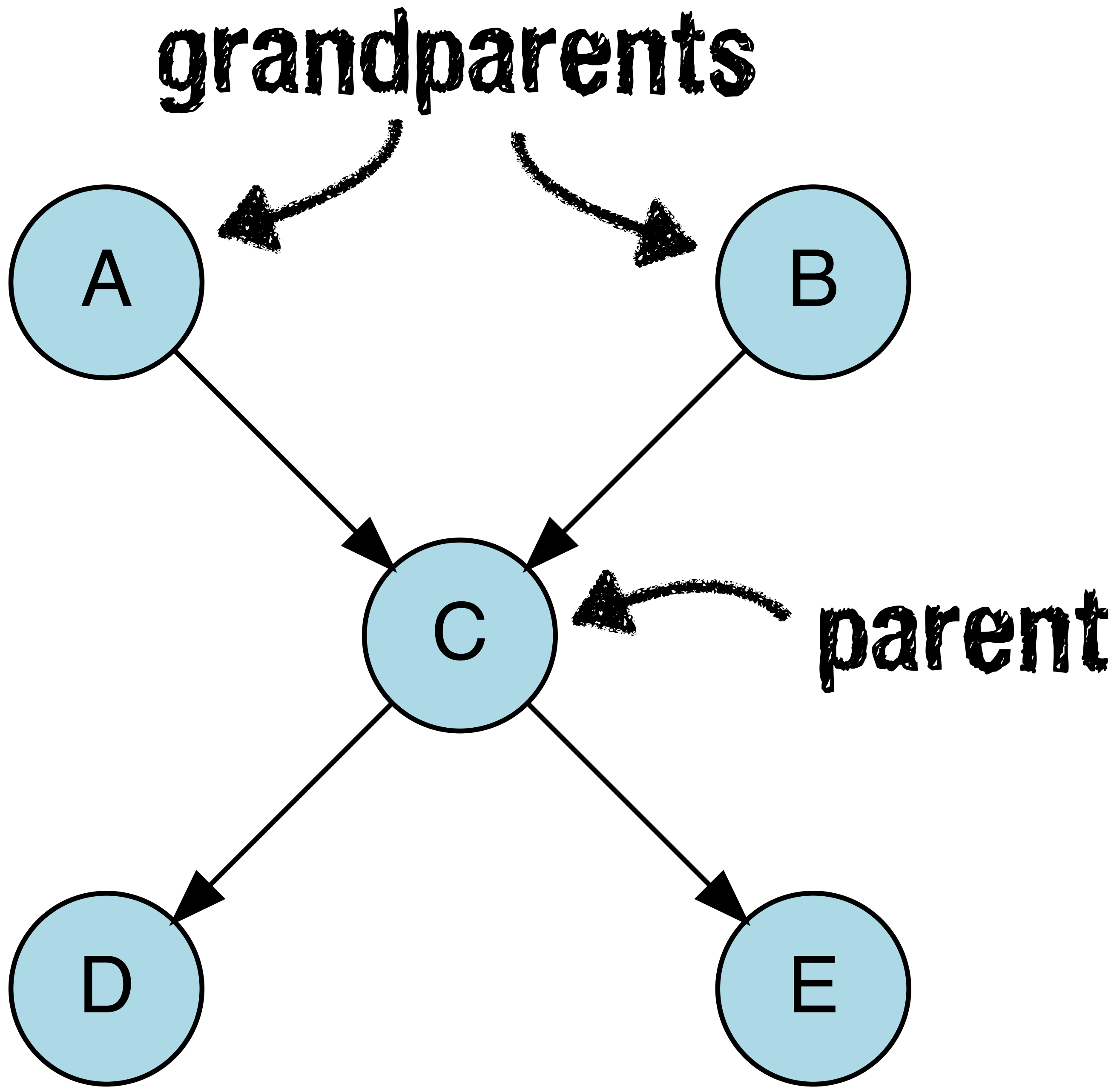
Are D and E independent?

$$p(D \mid E) = p(D) \;?$$



1. Draw the ancestral graph

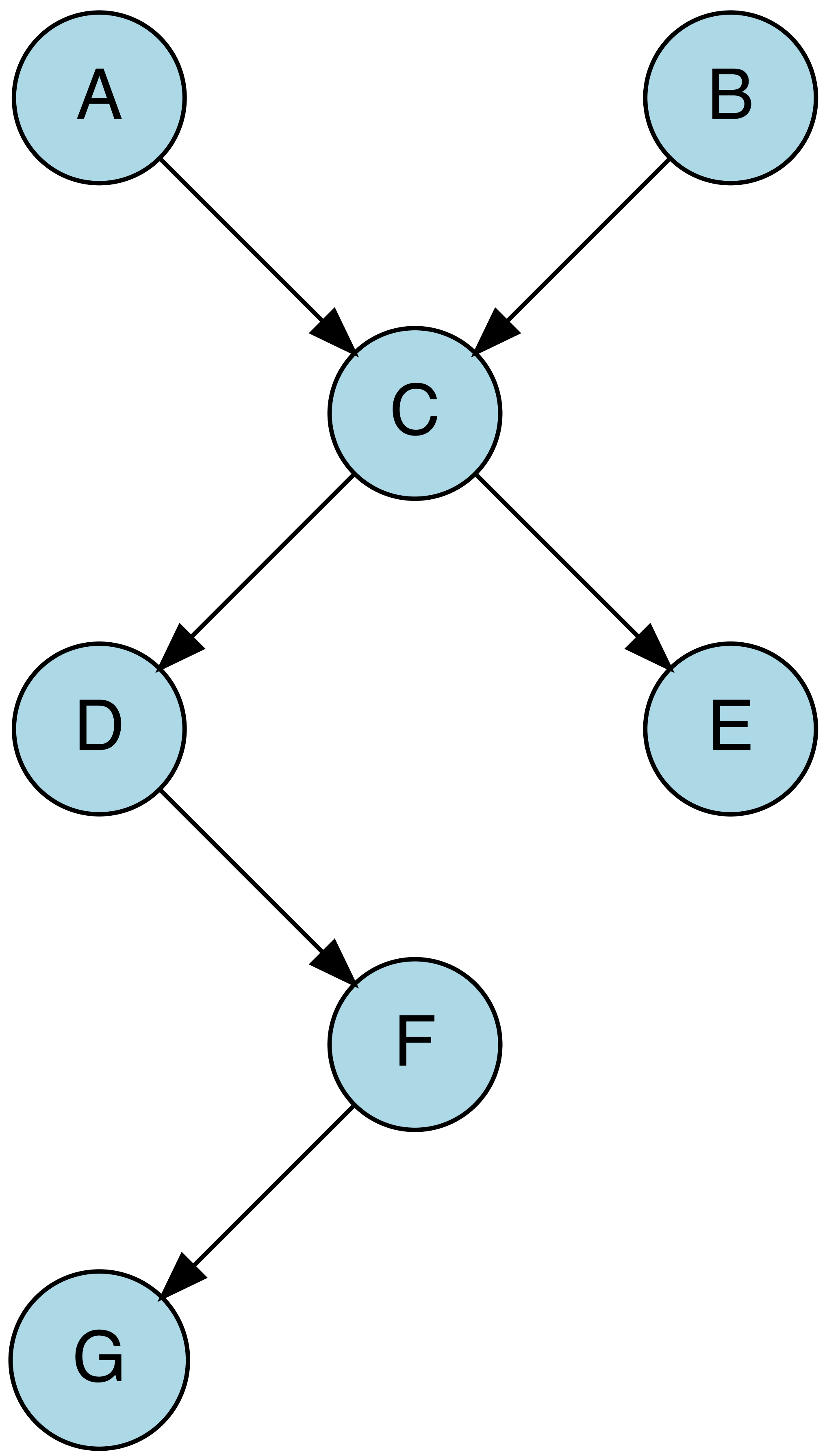
Construct the "ancestral graph" of all variables mentioned in the probability expression. This is a reduced version of the original net, consisting only of the variables mentioned and all of their ancestors (parents, parents' parents, etc.)



When should I control for variables?

Are D and E independent?

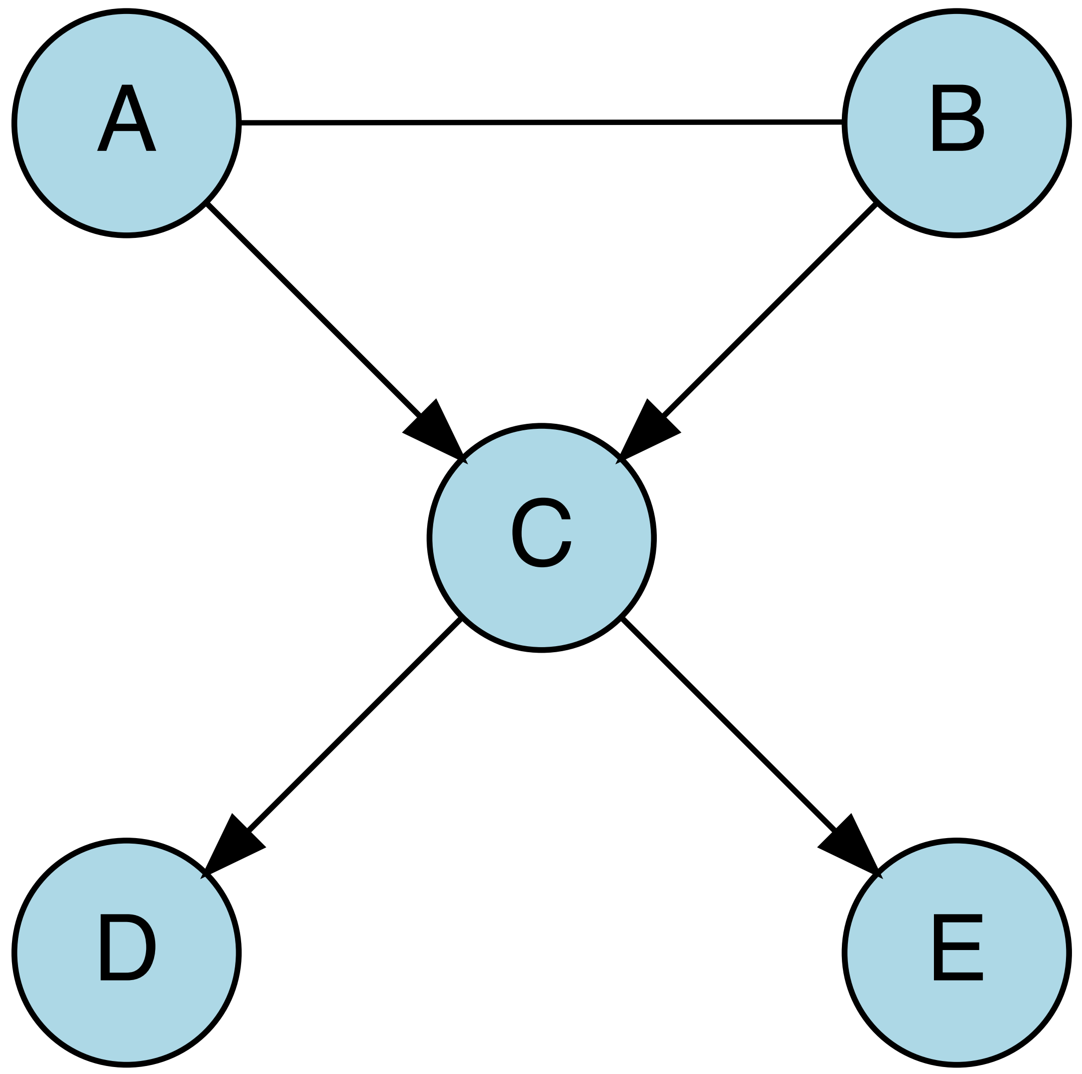
$$p(D\,|\,E) = p(D) \;\;?$$



2. "Moralize" the graph

let's get married!

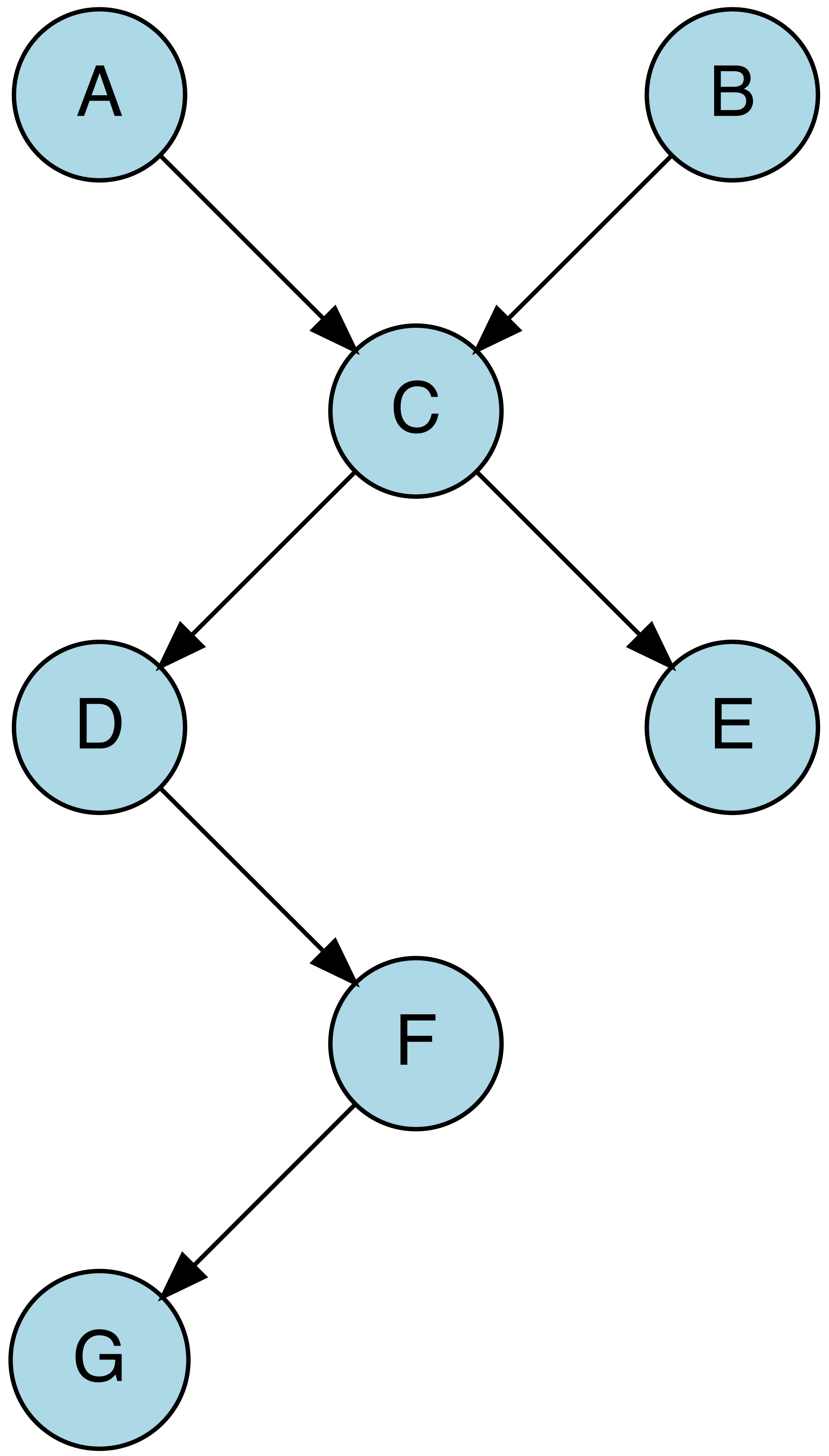
For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)



When should I control for variables?

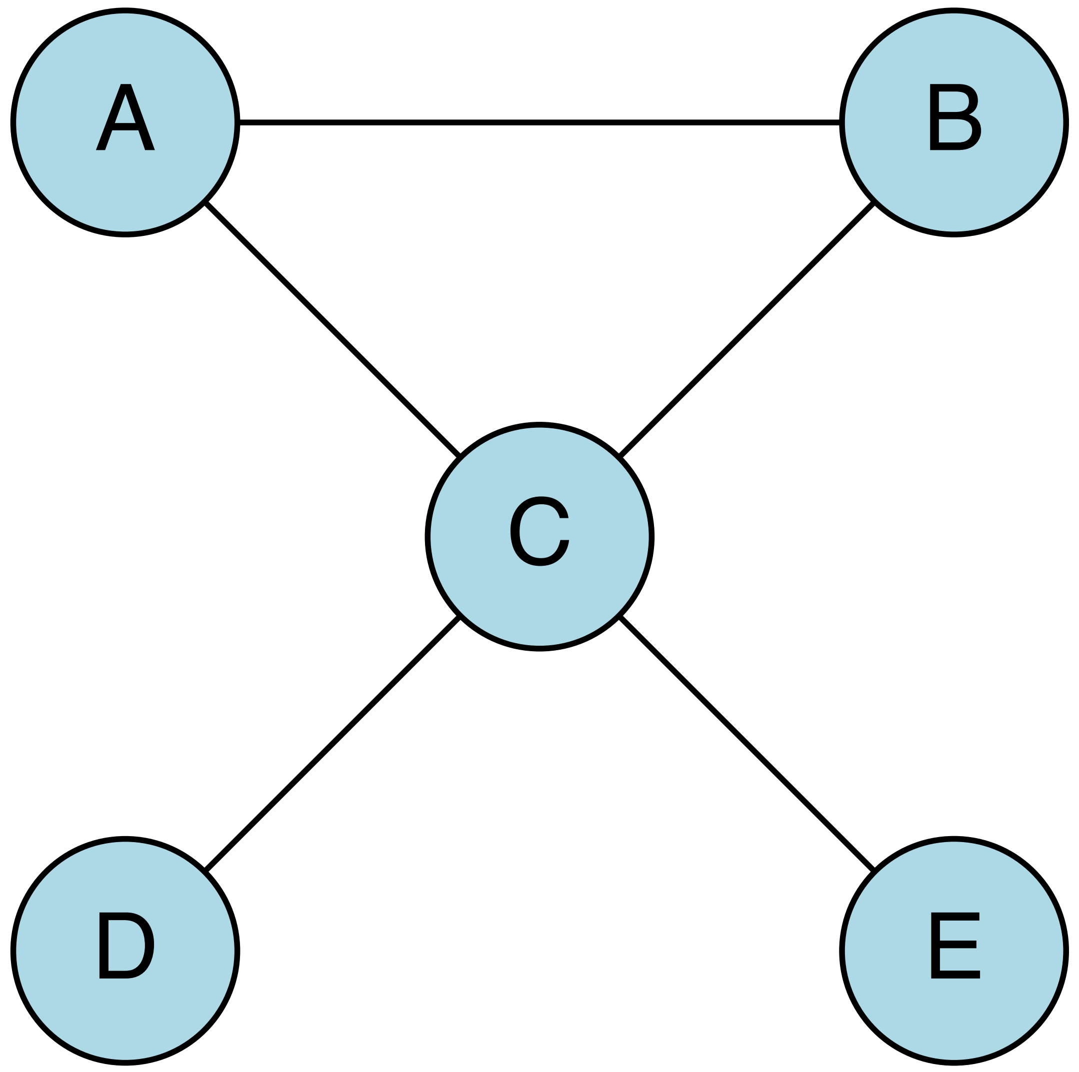
Are D and E independent?

$$p(D\,|\,E) = p(D)\ ?$$



3. "Disorient" the graph

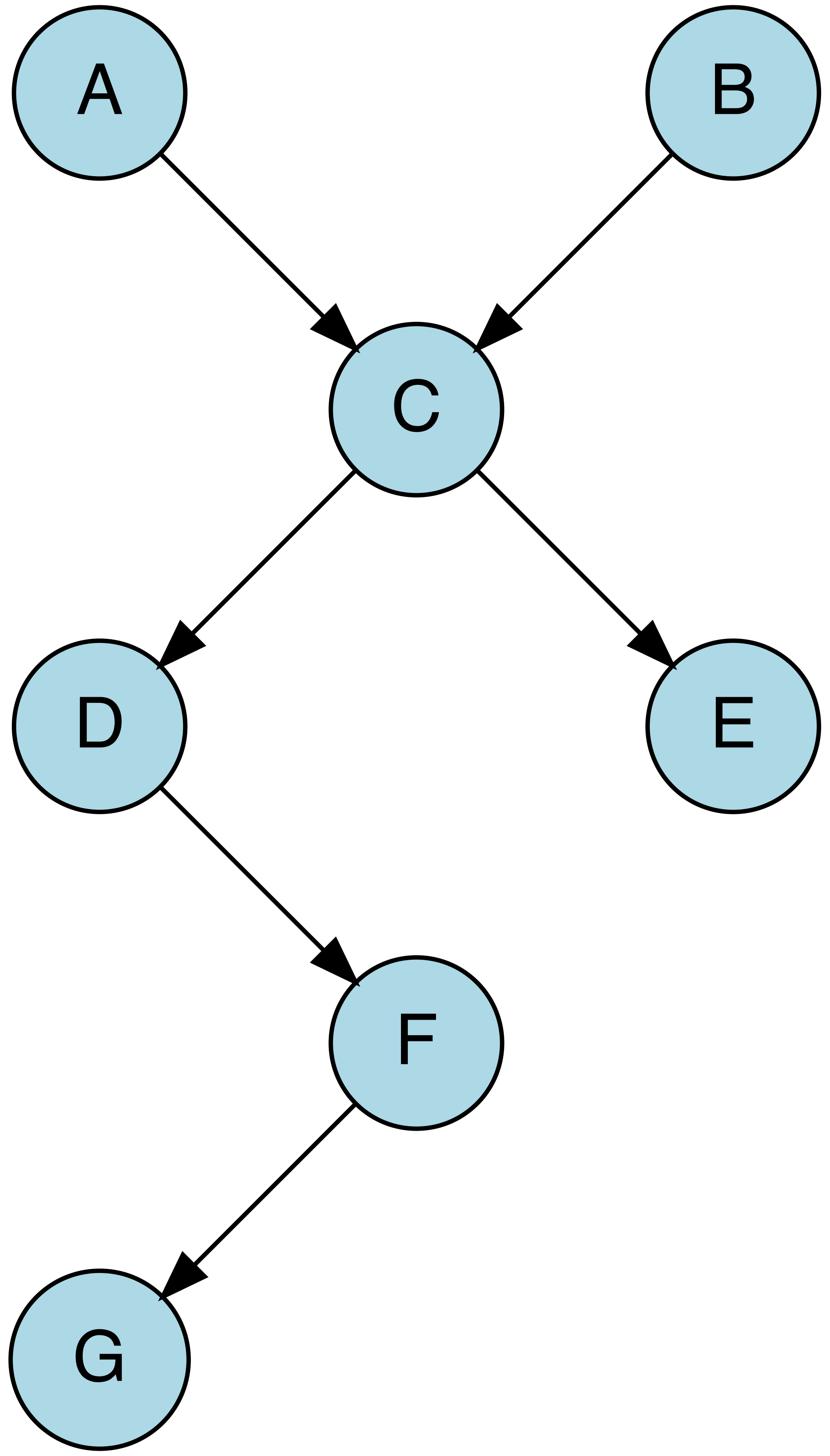
Replace arrows arrows with lines



When should I control for variables?

Are D and E independent?

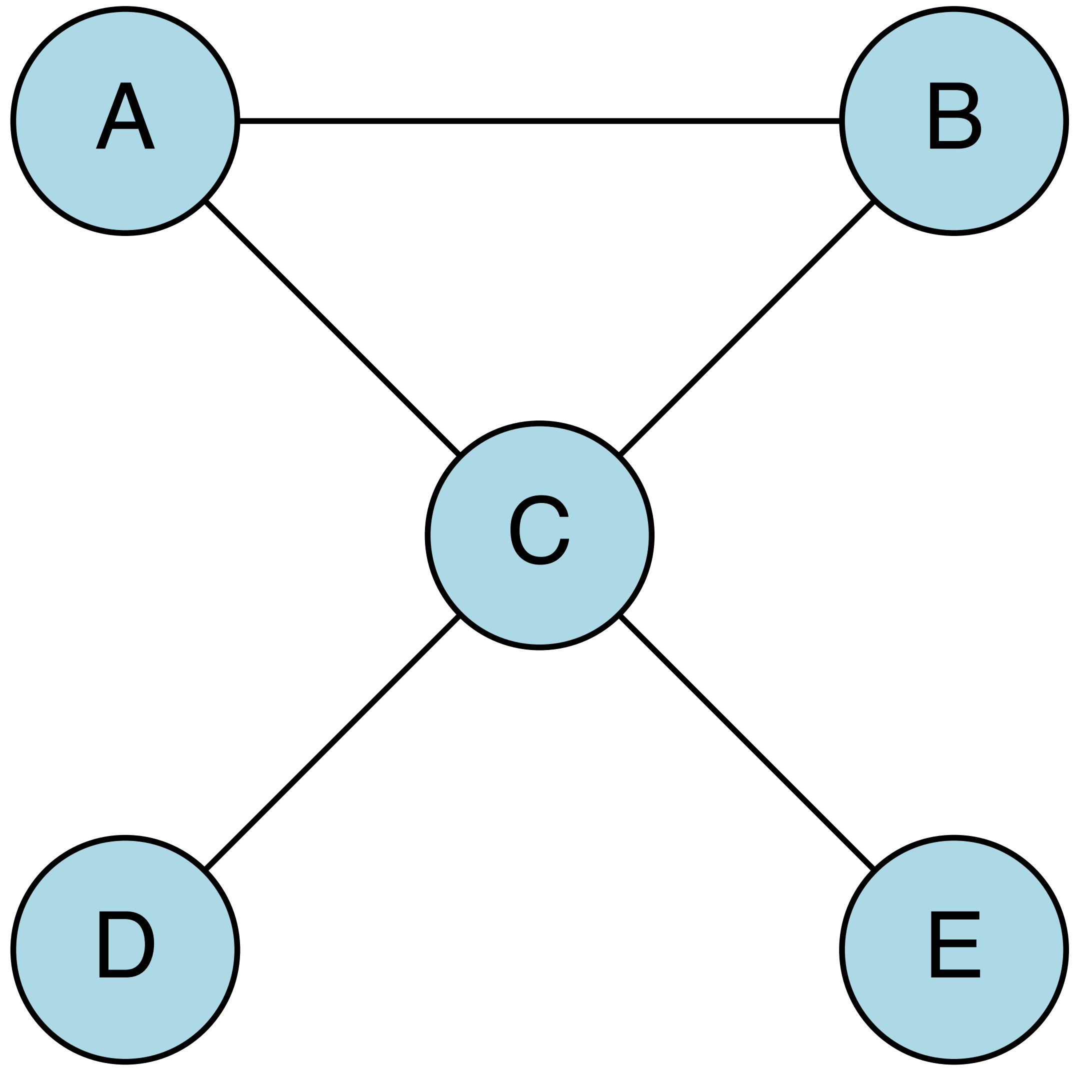
$$p(D \mid E) = p(D) \ ?$$



4. Delete the givens

Remove the variables that we condition on, as well as their edges

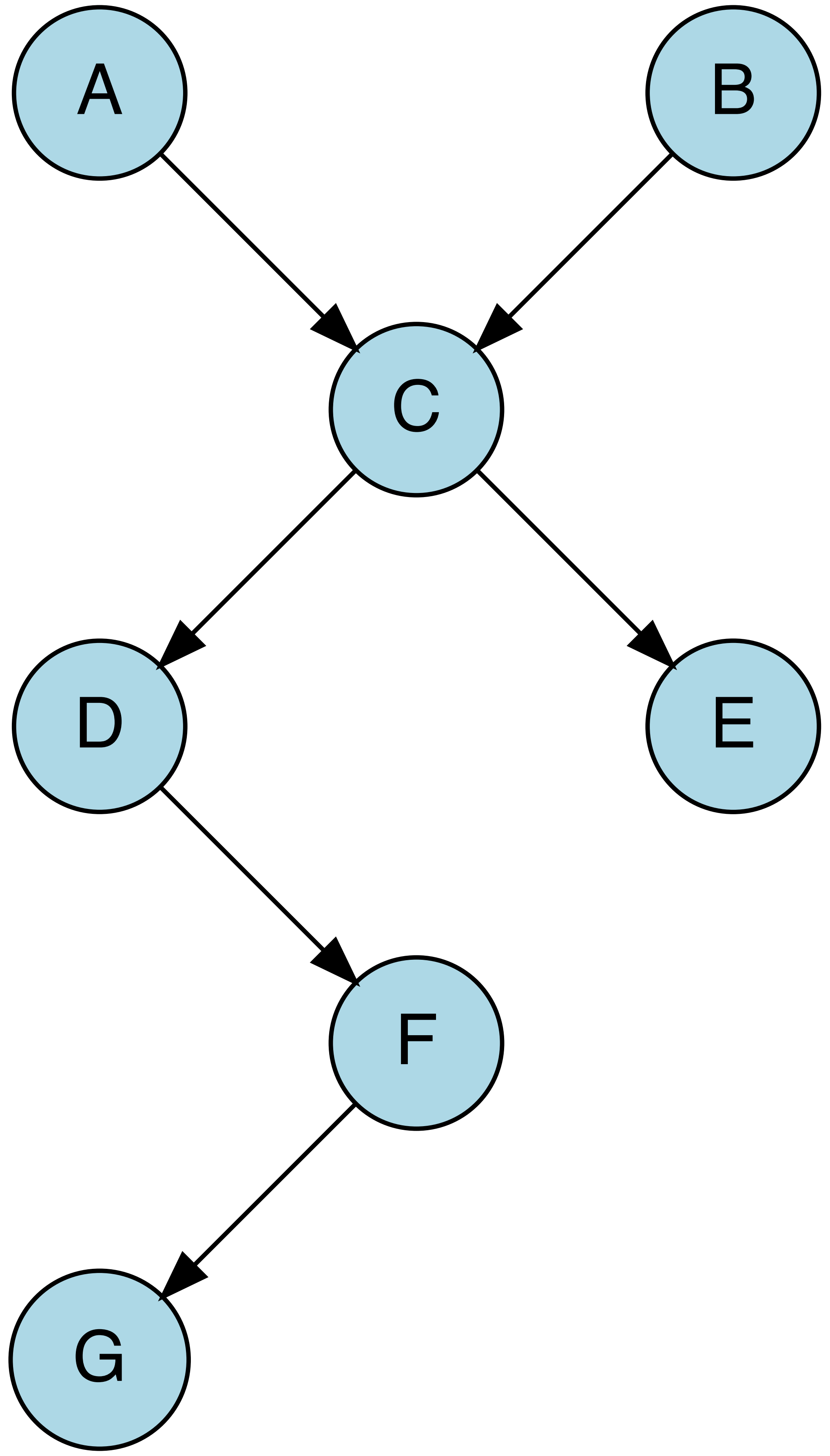
we didn't condition on anything, so there is nothing to delete



When should I control for variables?

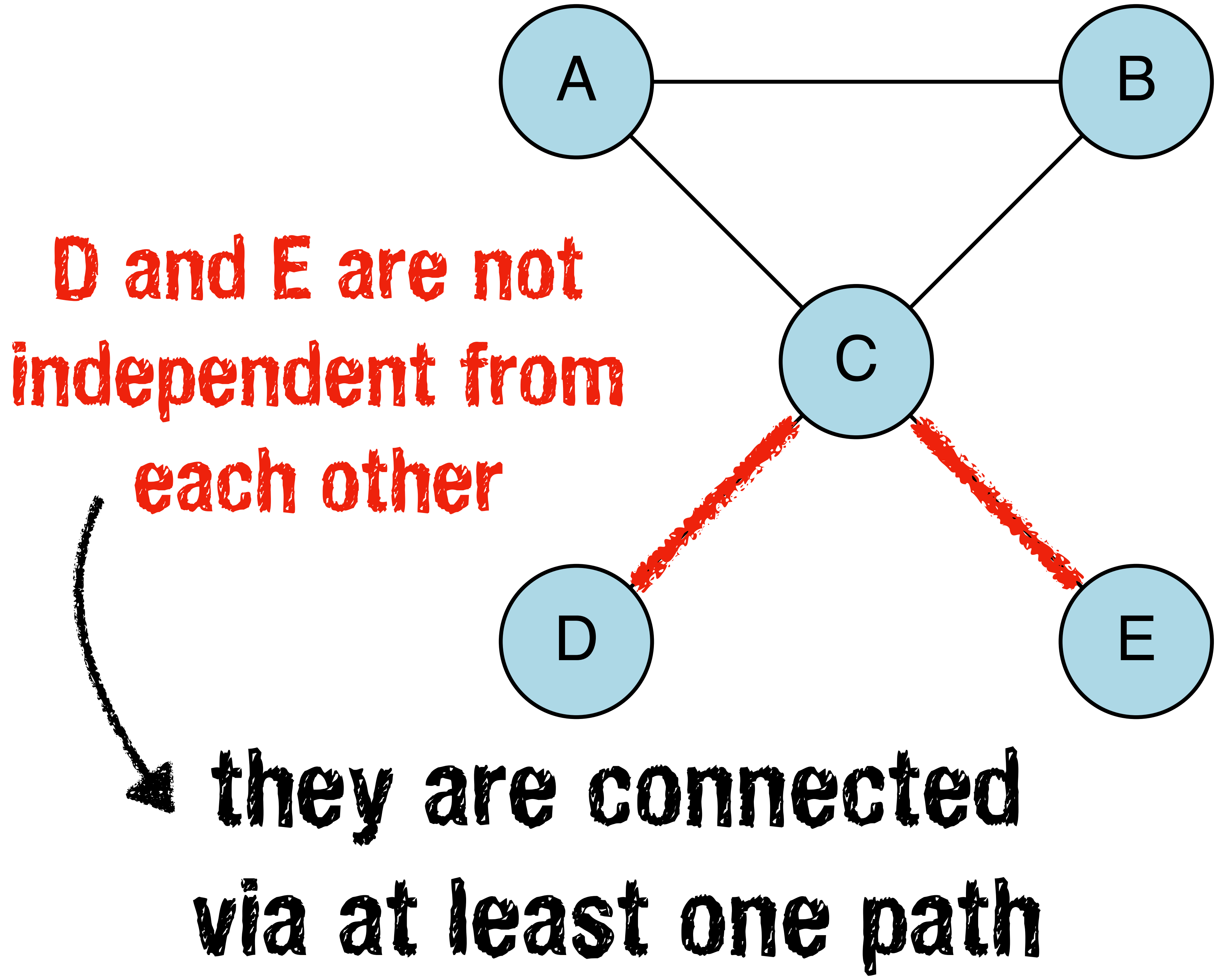
Are D and E independent?

$$p(D\,|\,E) = p(D) \text{ ?}$$



5. Read answer off the graph

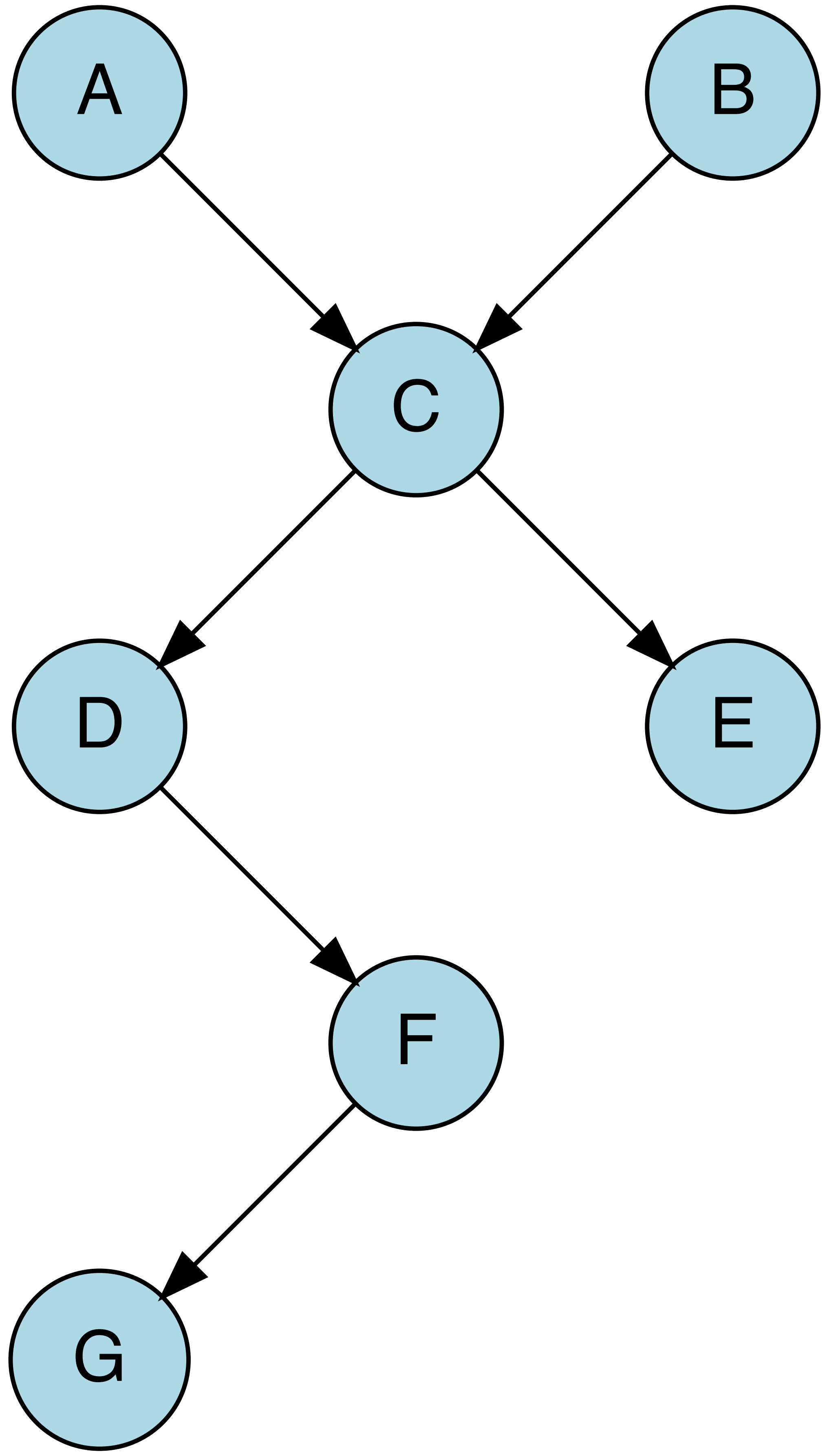
- if variables are **disconnected** they are independent
- if variables are connected (have a path between them) they are not guaranteed to be independent



When should I control for variables?

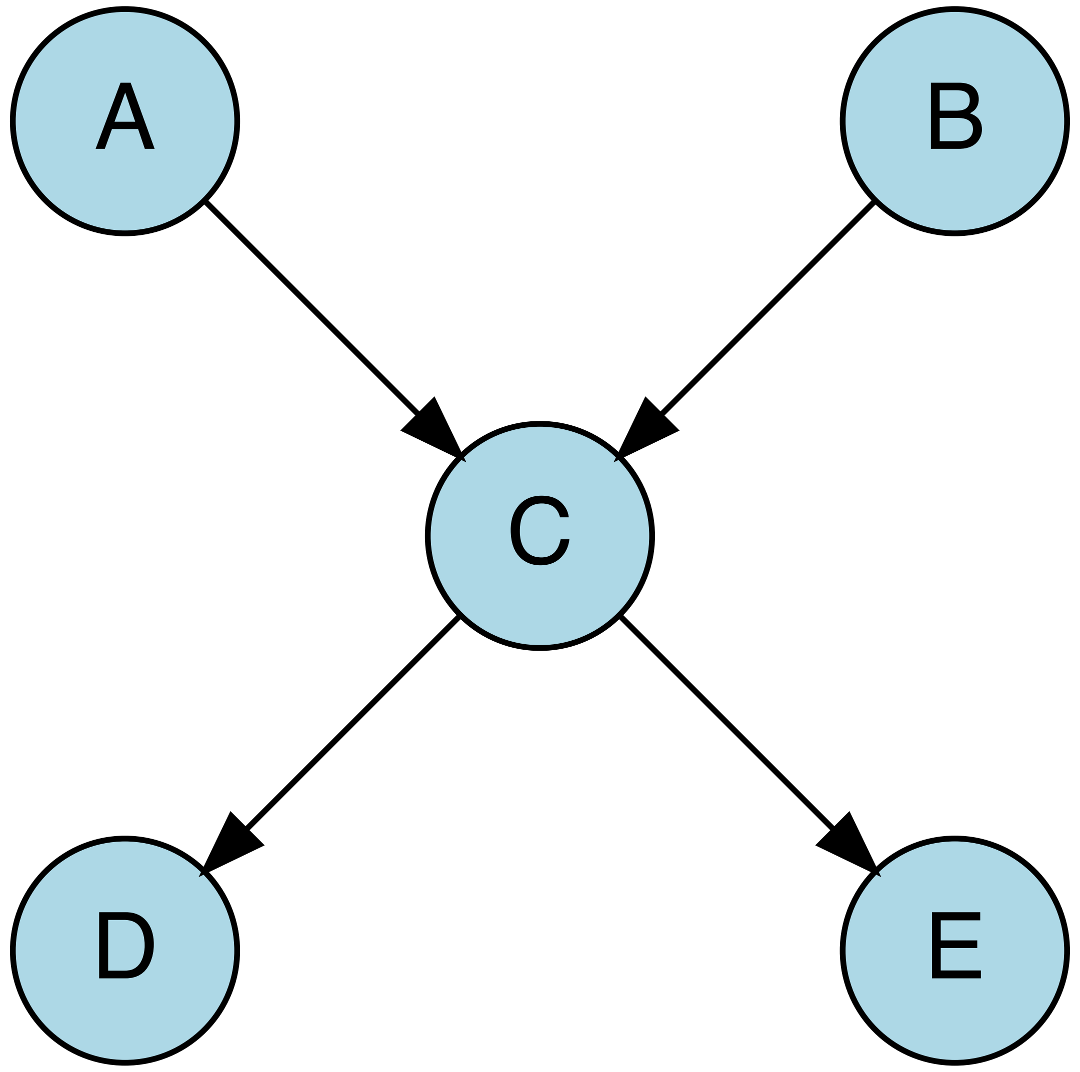
Are D and E independent, given C?

$$p(D \mid E, C) = p(D \mid C) \text{ ?}$$



1. Draw the ancestral graph

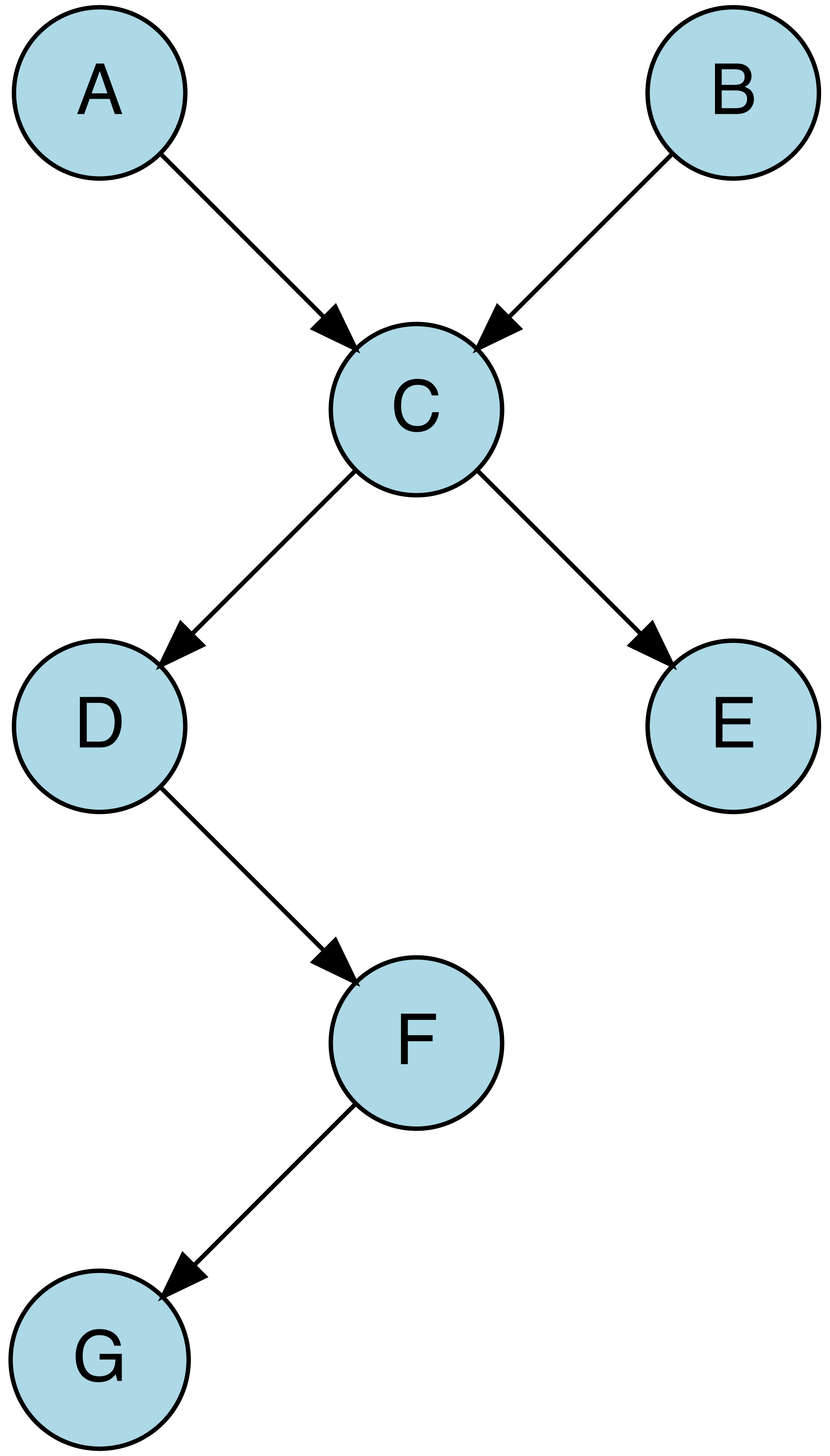
Construct the "ancestral graph" of all variables mentioned in the probability expression. This is a reduced version of the original net, consisting only of the variables mentioned and all of their ancestors (parents, parents' parents, etc.)



When should I control for variables?

Are D and E independent, given C?

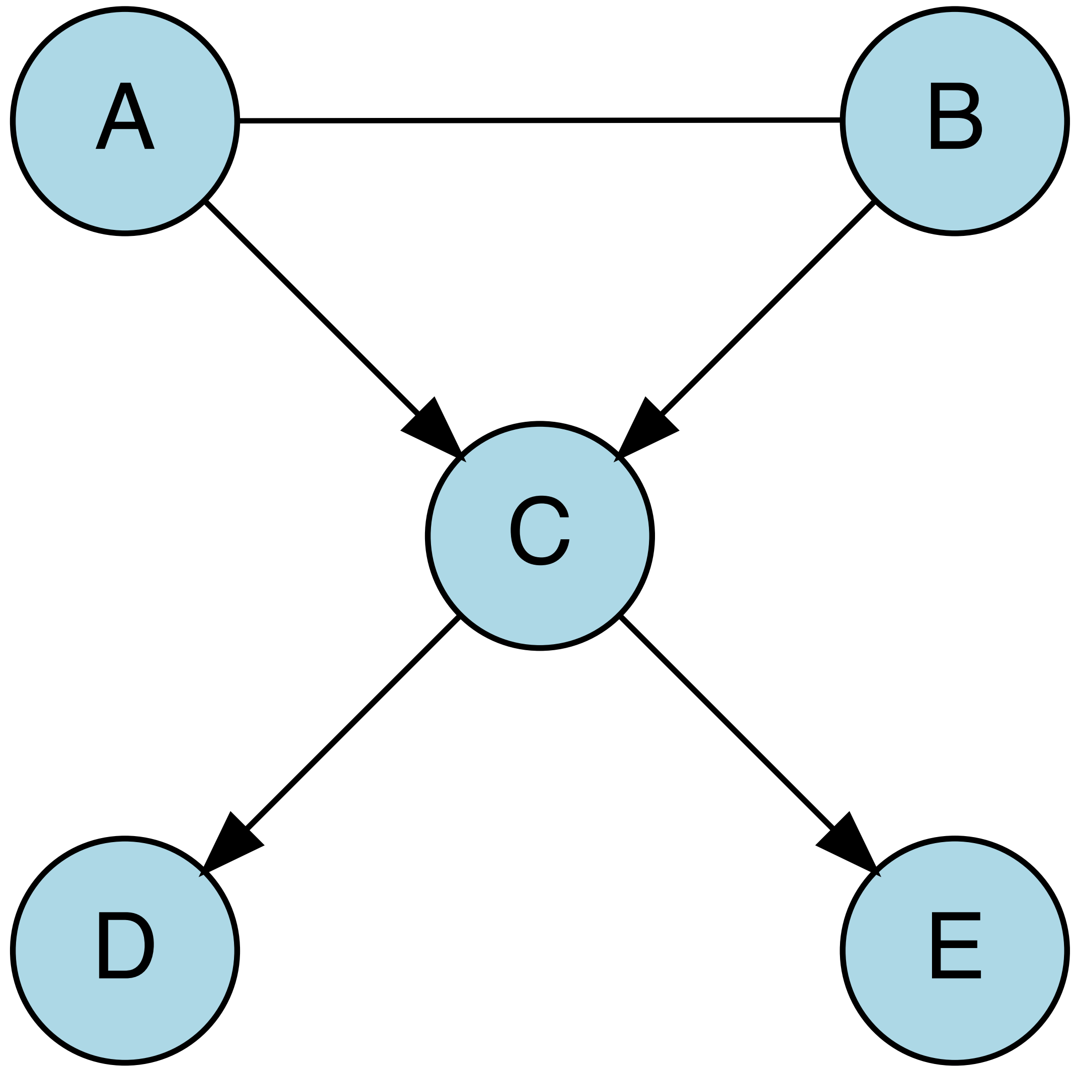
$$p(D\,|\,E, C) = p(D\,|\,C) \text{ ?}$$



2. "Moralize" the graph

let's get married!

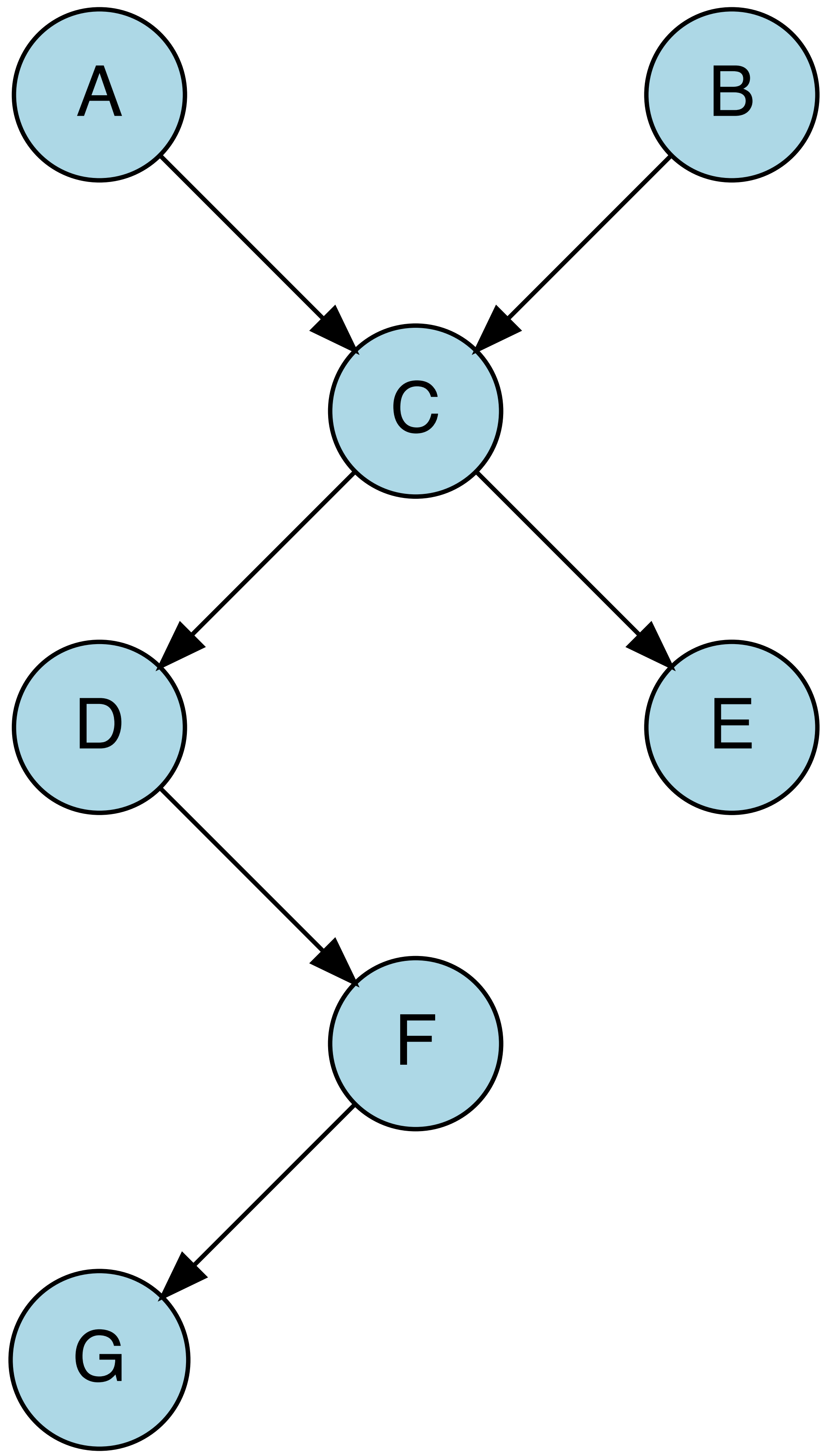
For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)



When should I control for variables?

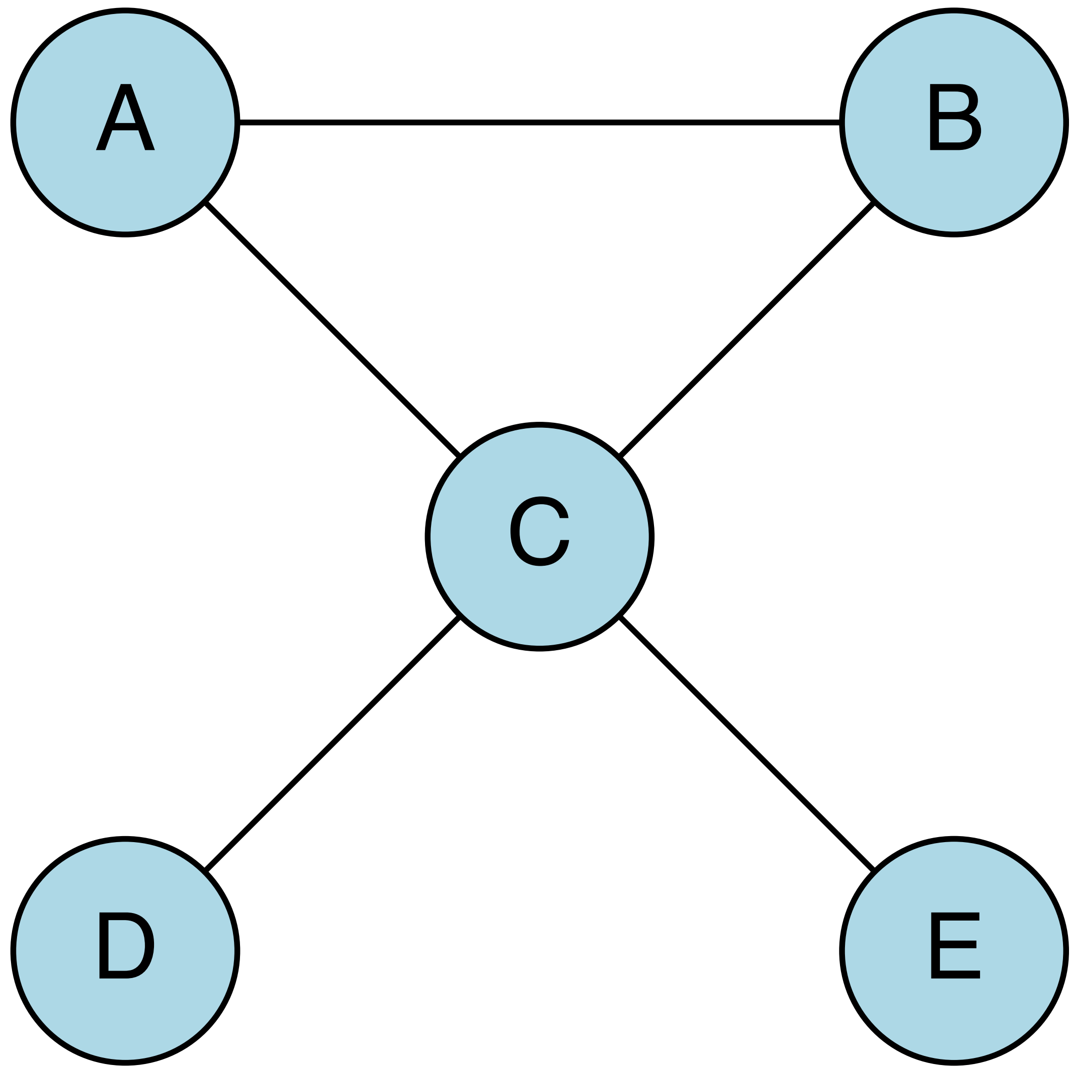
Are D and E independent, given C?

$$p(D \mid E, C) = p(D \mid C) \ ?$$



3. "Disorient" the graph

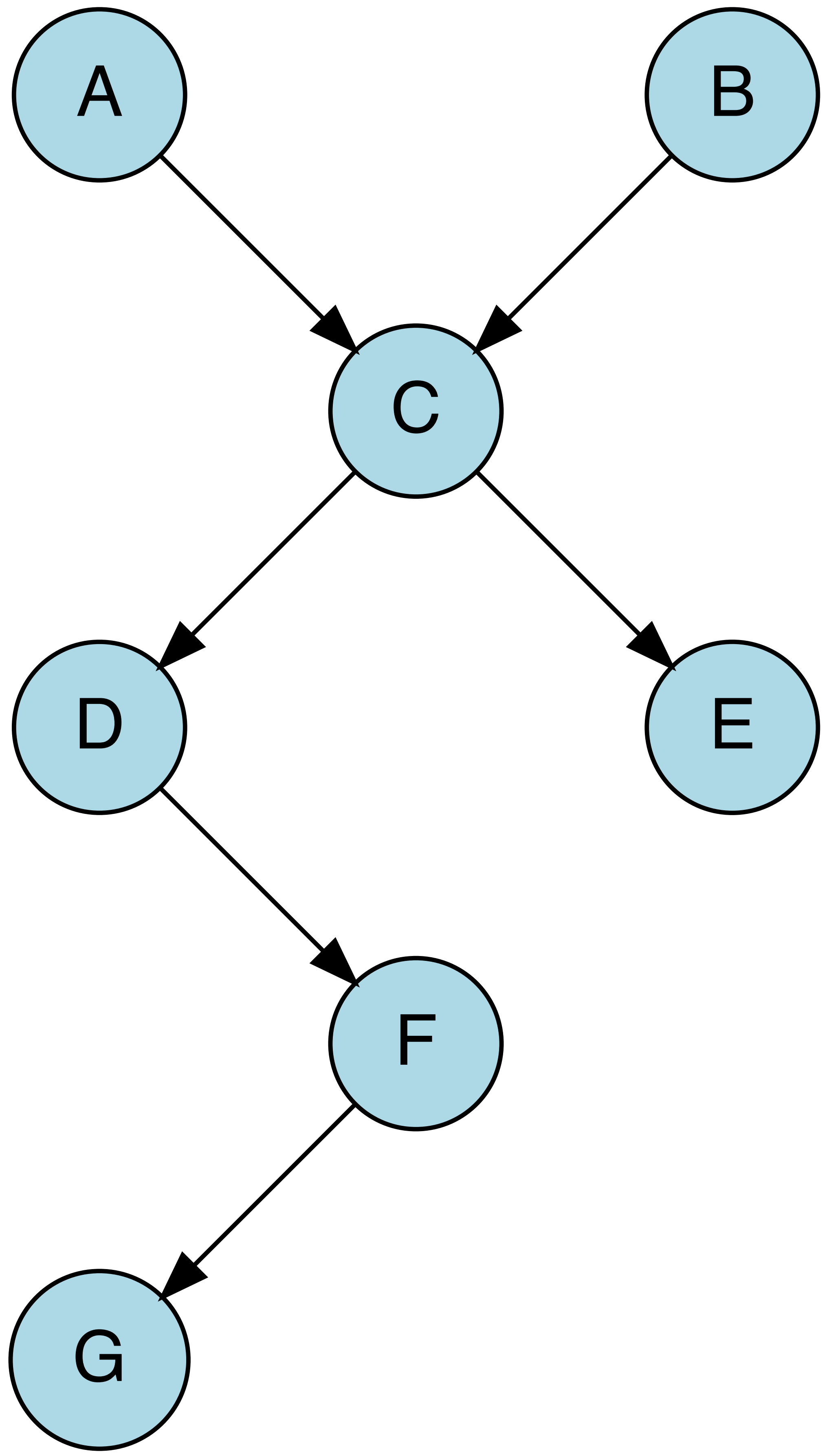
Replace arrows arrows with lines



When should I control for variables?

Are D and E independent, given C?

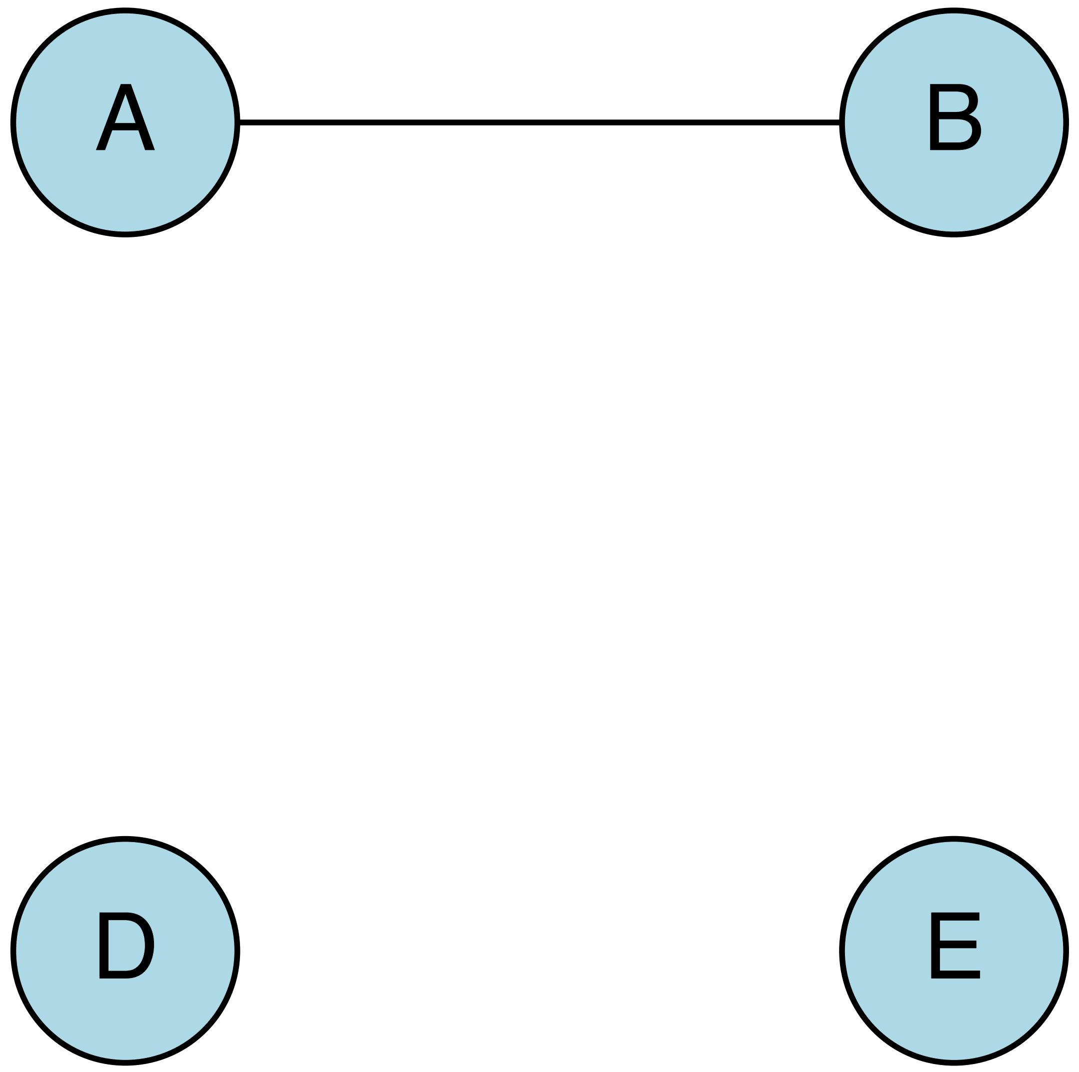
$$p(D \mid E, C) = p(D \mid C) \; ?$$



4. Delete the givens

Remove the variables that we condition on, as well as their edges

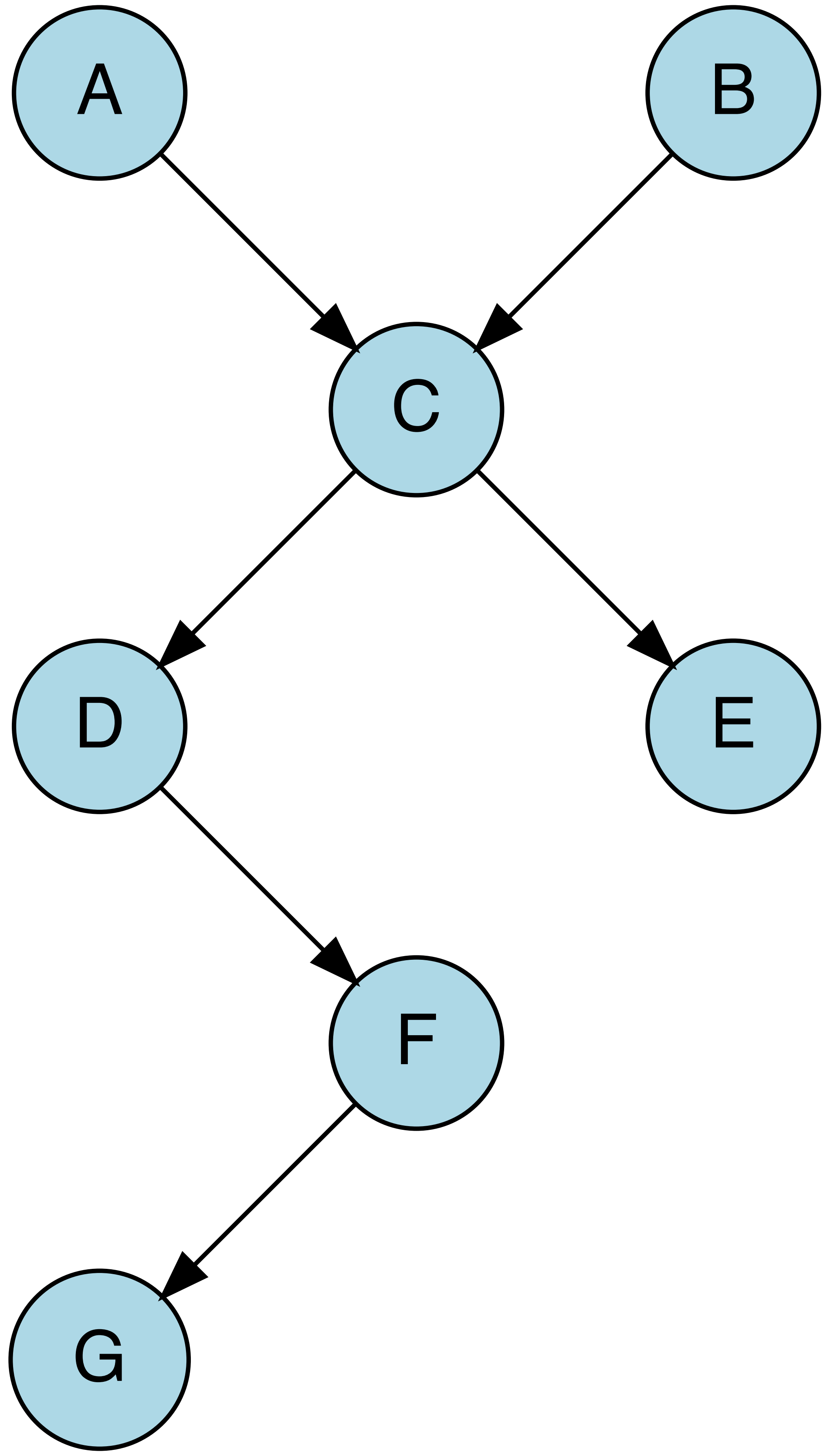
we conditioned on C!



When should I control for variables?

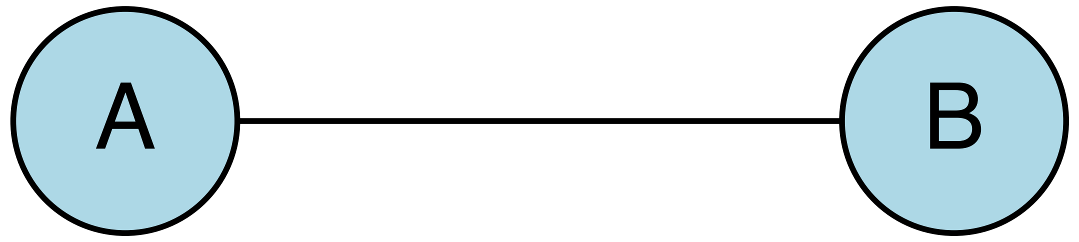
Are D and E independent, given C?

$$p(D\,|\,E, C) = p(D\,|\,C) \text{ ?}$$



5. Read answer off the graph

- if variables are **disconnected** they are independent
- if variables are connected (have a path between them) they are not guaranteed to be independent



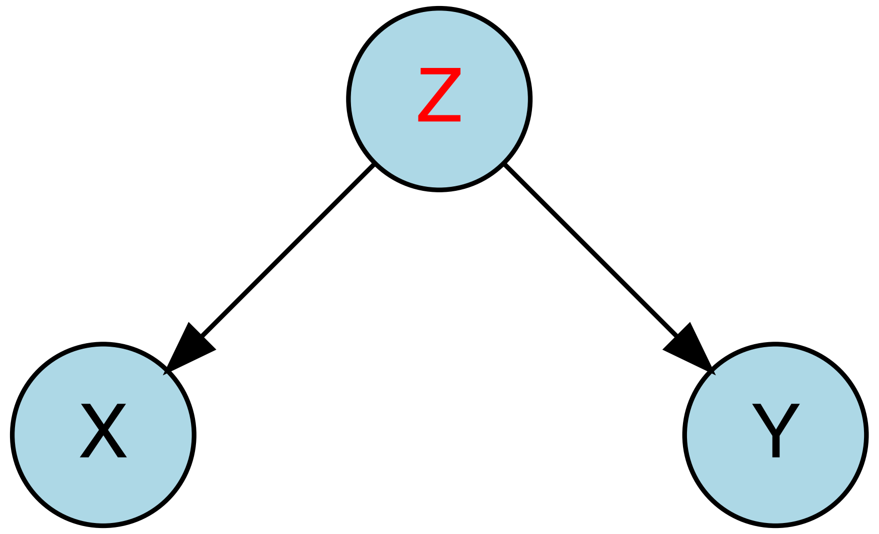
D and E are independent from each other conditioned on C

they aren't connected via a path

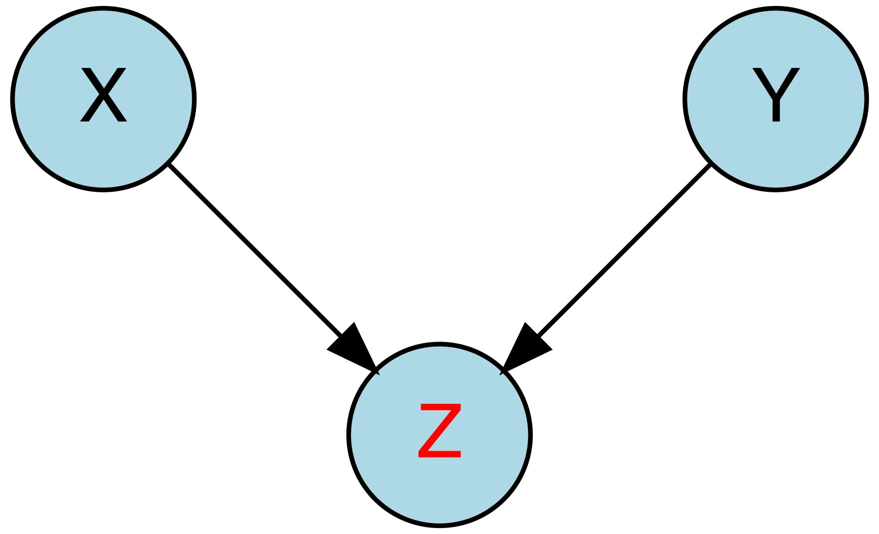
So what?

Patterns of inference

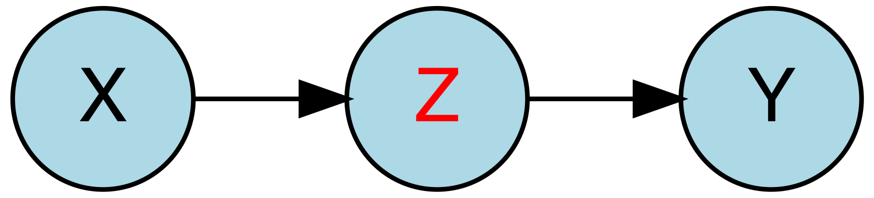
We want to estimate the (causal) relationship between `X` and `Y`



common cause



common effect



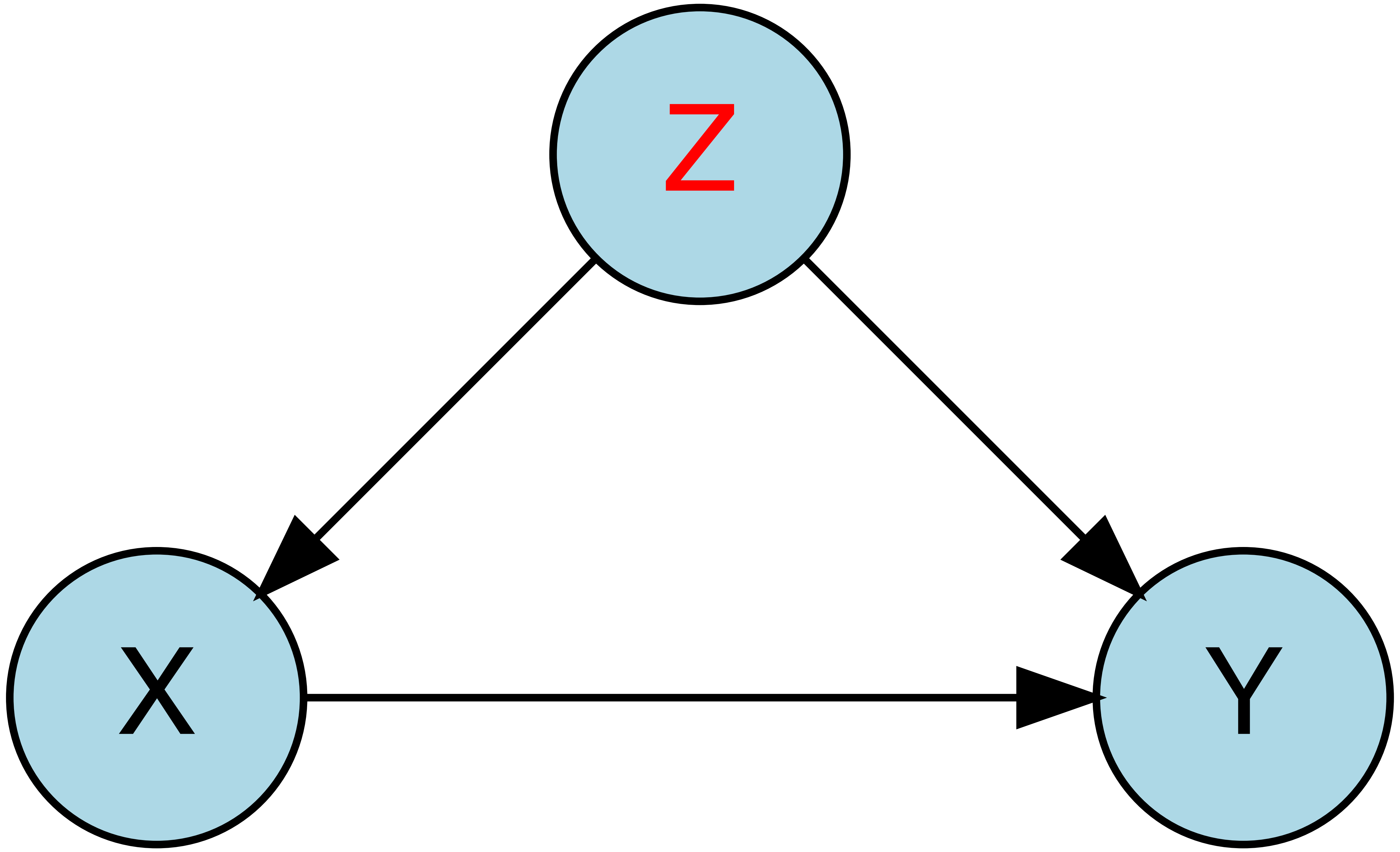
causal chain

by controlling for `Z` we hope to get a better estimate of the relationship between `X` and `Y`

d-separation helps us tell apart **good controls** from **bad controls**

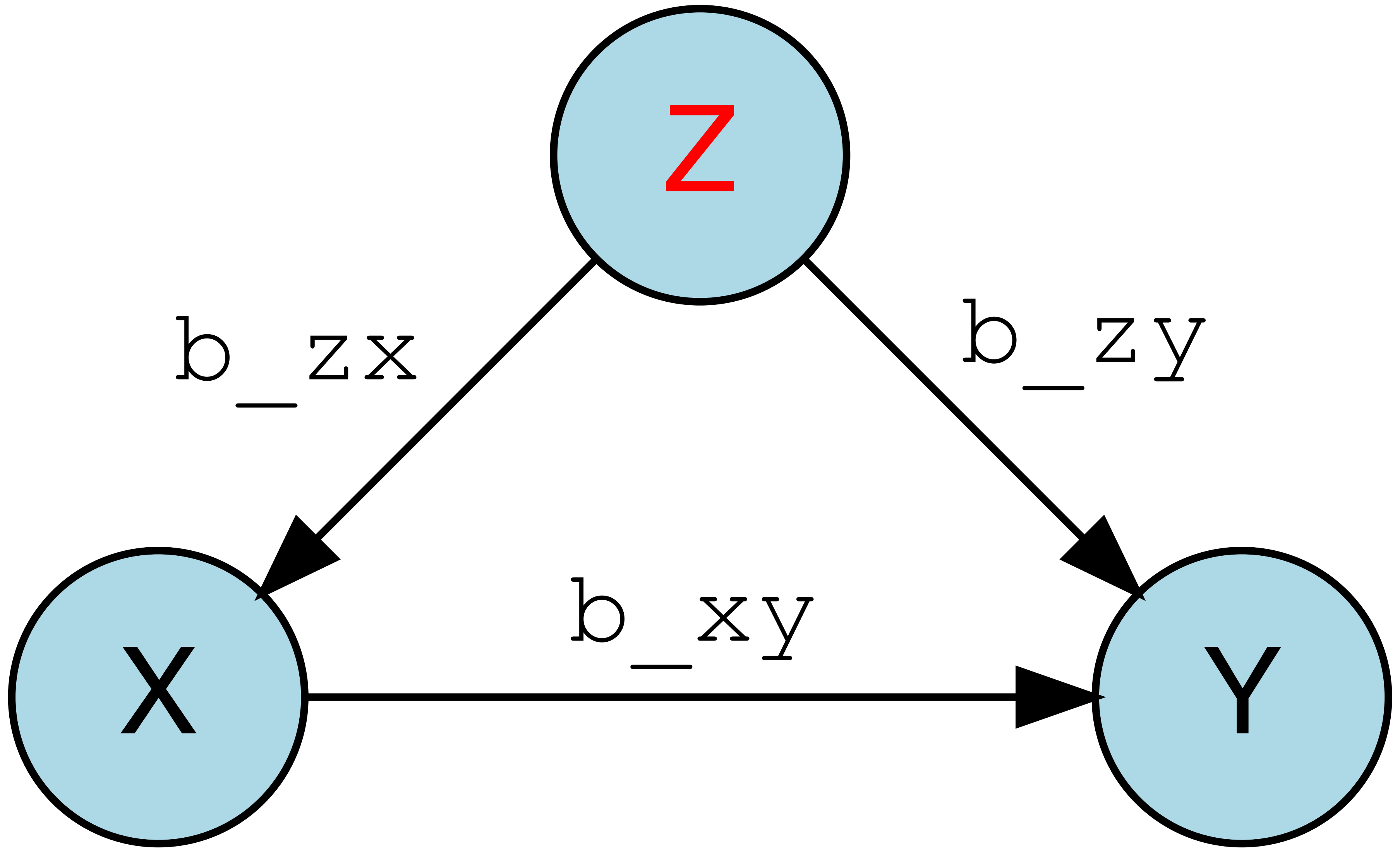
When should I control for variables?

I want to estimate the effect that X has on Y



Is Z a **good** or a **bad** control here?

When should I control for variables?




```r
set.seed(1)

n = 1000
b_zx = 2
b_xy = 2
b_zy = 2
sd = 1

df = tibble(z = rnorm(n = n, sd = sd),
            x = b_zx * z + rnorm(n = n, sd = sd),
            y = b_zy * z + b_xy * x + rnorm(n = n, sd = sd))
```

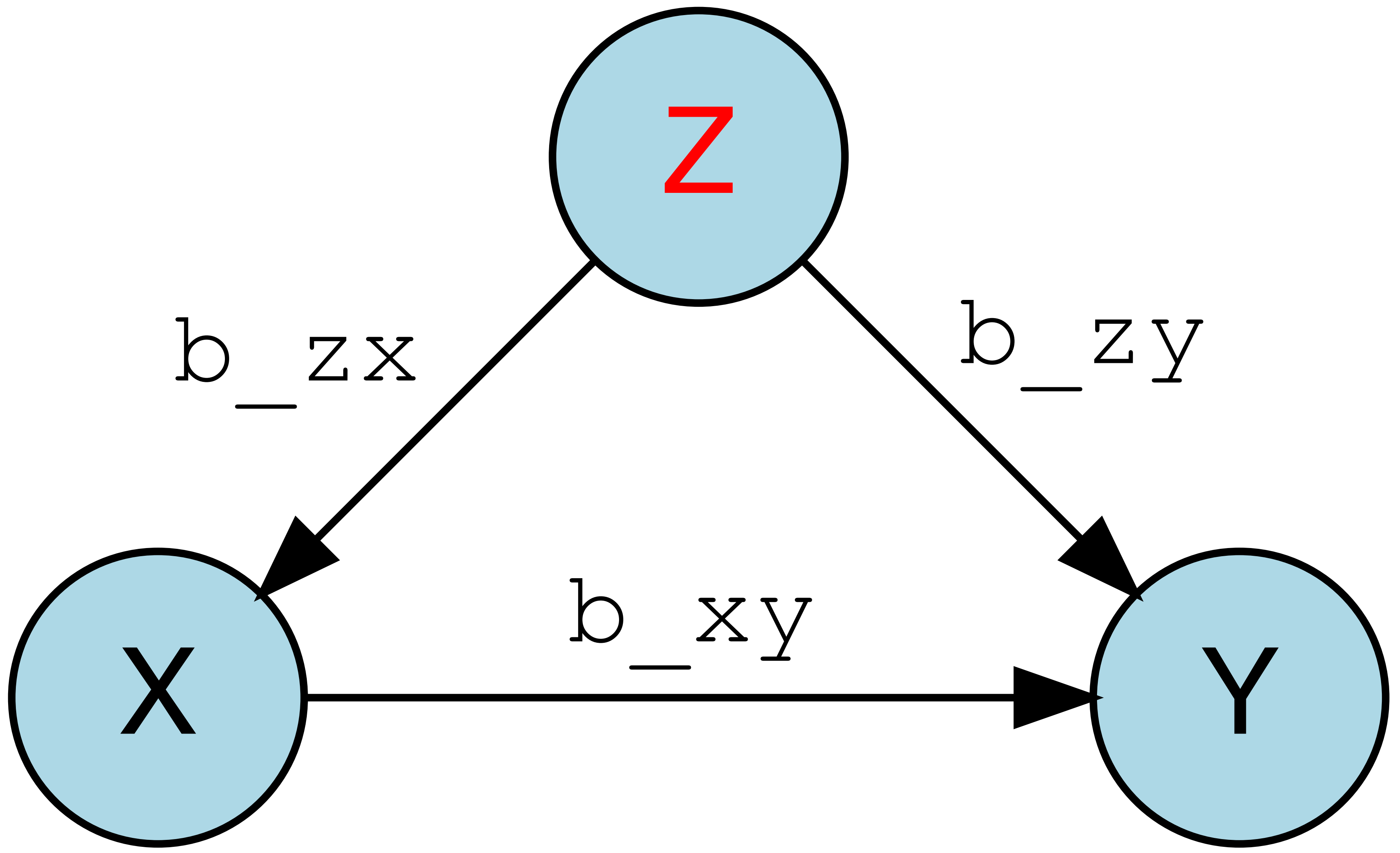
overestimating X's effect on Y

$$Y = b_0 + b_1 \cdot X + e$$

```r
# without control
lm(formula = y ~ 1 + x,
   data = df) %>%
  summary()
```

```
Call:
lm(formula = y ~ x, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6011 -0.9270 -0.0506  0.9711  4.0454

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02449    0.04389   0.558    0.577
x            2.82092    0.01890 149.225   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.388 on 998 degrees of freedom
Multiple R-squared:  0.9571,	Adjusted R-squared:  0.9571
F-statistic: 2.227e+04 on 1 and 998 DF,  p-value: < 2.2e-16
```

When should I control for variables?




```r
set.seed(1)

n = 1000
b_zx = 2
b_xy = 2
b_zy = 2
sd = 1

df = tibble(z = rnorm(n = n, sd = sd),
            x = b_zx * z + rnorm(n = n, sd = sd),
            y = b_zy * z + b_xy * x + rnorm(n = n, sd = sd))
```

$$Y = b_0 + b_1 \cdot X + b_2 \cdot Z + e$$

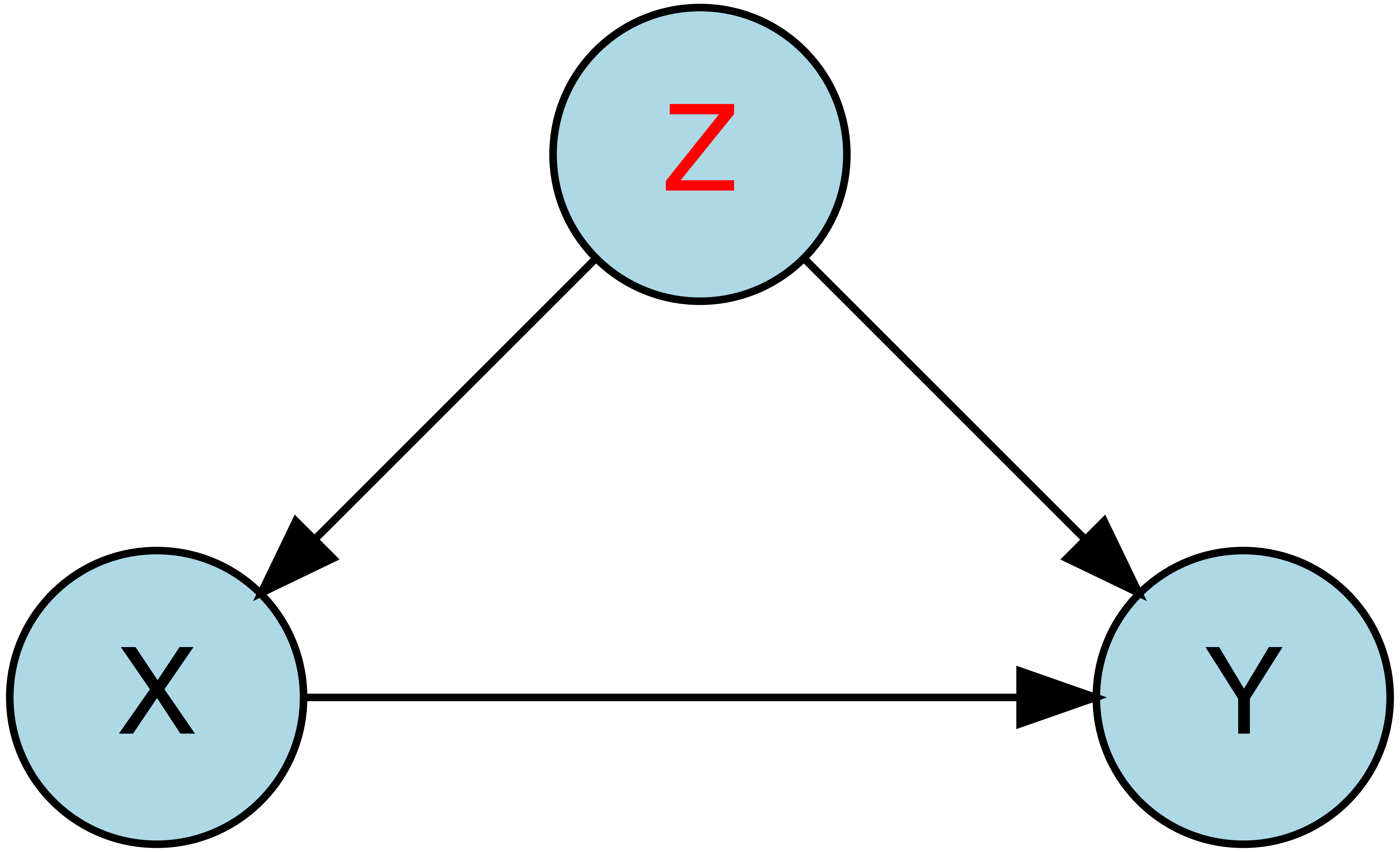
```r
# with control
lm(formula = y ~ 1 + x + z,
   data = df) %>%
  summary()
```

accurate estimate of X's effect on Y

```
Call:
lm(formula = y ~ x + z, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6151 -0.6564 -0.0223  0.6815  2.8132 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.01624    0.03260   0.498    0.618    
x            2.02202    0.03135  64.489   <2e-16 ***
z            2.00501    0.07036  28.497   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.031 on 997 degrees of freedom
Multiple R-squared:  0.9764,	Adjusted R-squared:  0.9763 
F-statistic: 2.059e+04 on 2 and 997 DF,  p-value: < 2.2e-16
```

When should I control for variables?

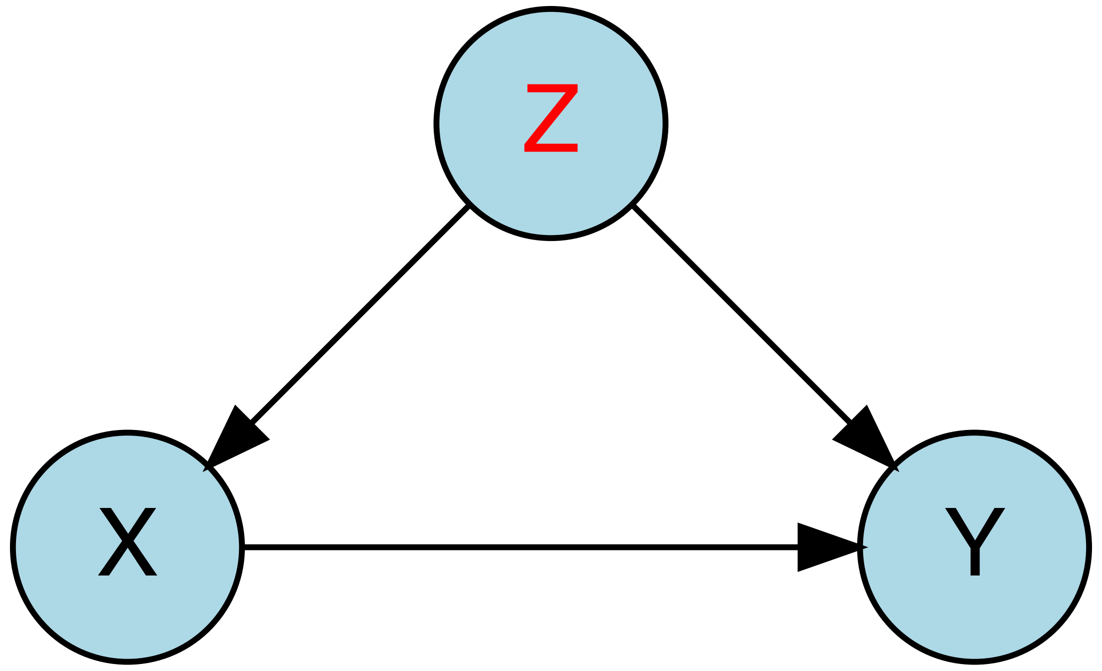
I want to estimate the effect that X has on Y



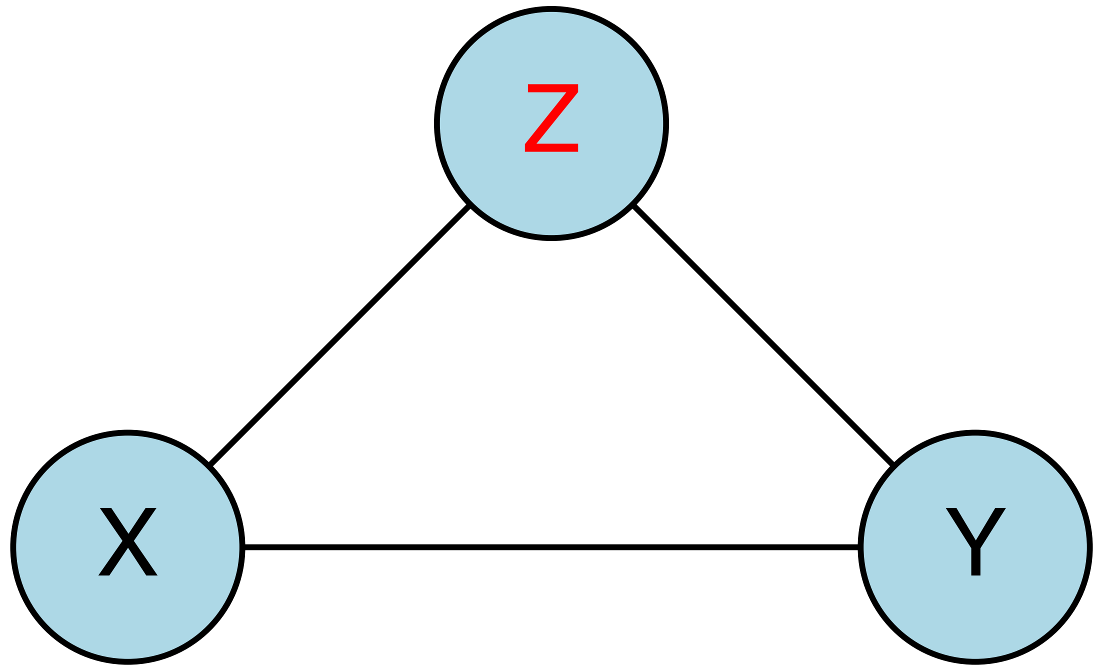
Z is a **good** control here!

When should I control for variables?

I want to estimate the effect that X has on Y

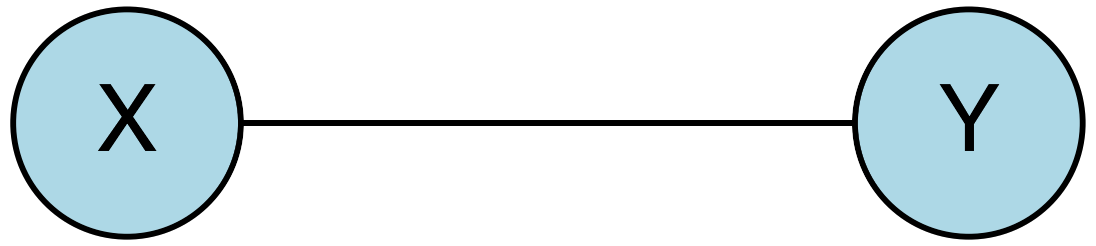


not controlling for Z



multiple paths

controlling for Z

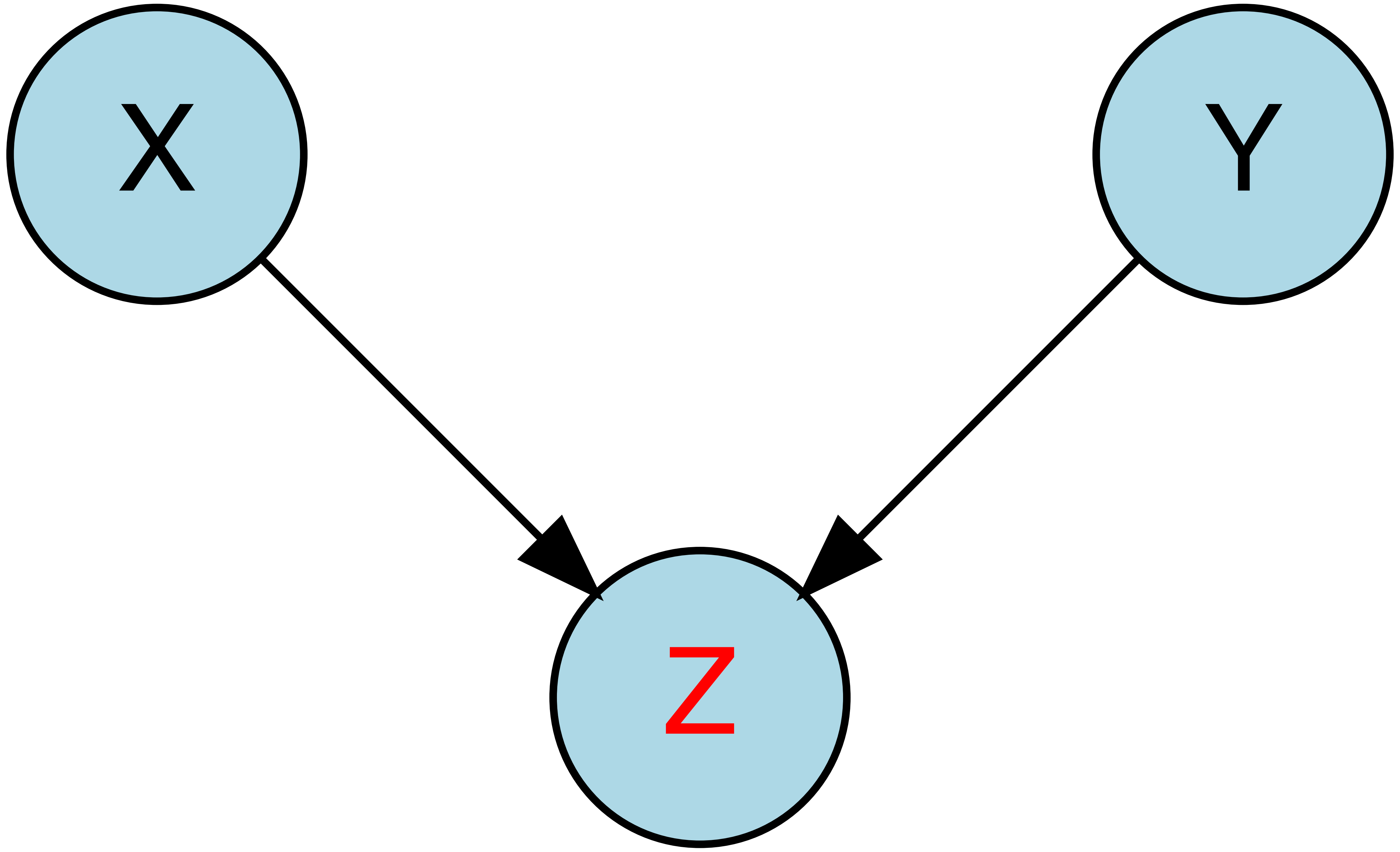


one path

Z is a **good** control here!

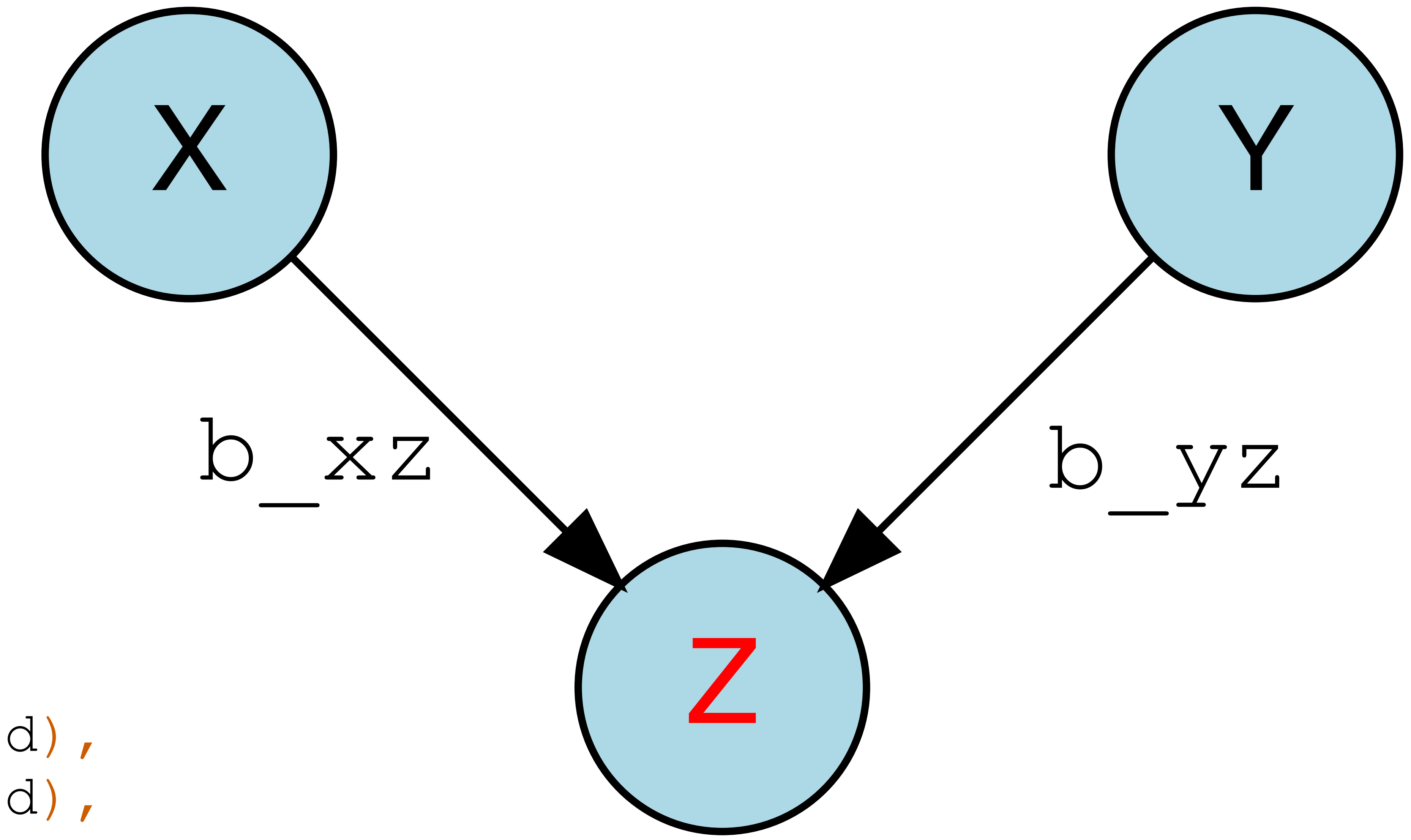
When should I control for variables?

I want to estimate the effect that X has on Y



Is Z a **good** or a **bad** control here?

When should I control for variables?




```r
set.seed(1)

n = 1000
b_xz = 2
b_yz = 2
sd = 1

df = tibble(x = rnorm(n = n, sd = sd),
            y = rnorm(n = n, sd = sd),
            z = x * b_xz + y * b_yz + rnorm(n = n, sd = sd))
```

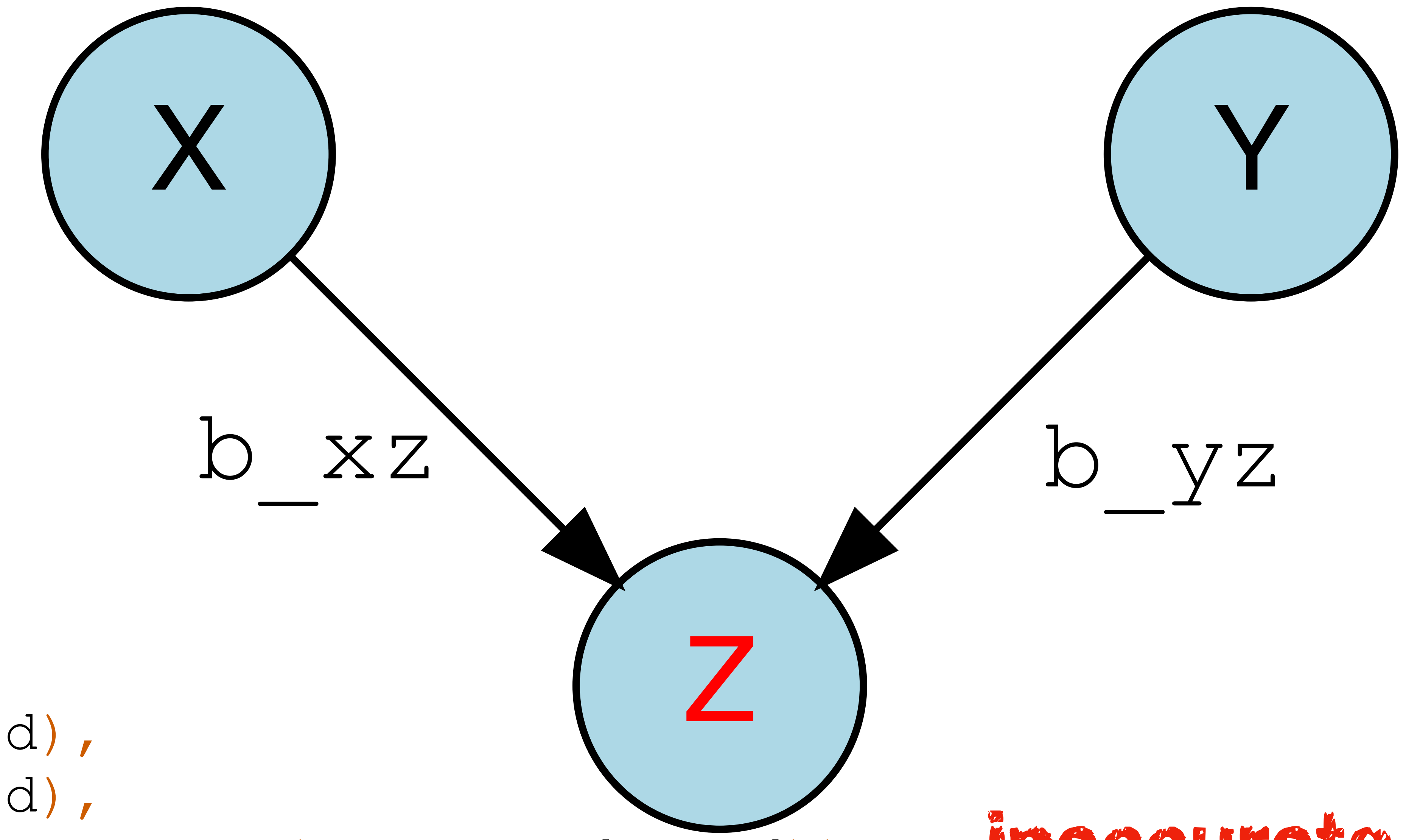
accurate estimate of X's effect on Y

$$Y = b_0 + b_1 \cdot X + e$$

```r
# without control
lm(formula = y ~ 1 + x,
   data = df) %>%
  summary()
```

```
Call:
lm(formula = y ~ x, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2484 -0.6720 -0.0138  0.7554  3.6443 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.016187   0.032905  -0.492    0.623
x            0.006433   0.031809   0.202    0.840

Residual standard error: 1.04 on 998 degrees of freedom
Multiple R-squared:  4.098e-05,   Adjusted R-squared:  -0.000961 
F-statistic: 0.0409 on 1 and 998 DF,  p-value: 0.8398
```

When should I control for variables?




```r
set.seed(1)

n = 1000
b_xz = 2
b_yz = 2
sd = 1

df = tibble(x = rnorm(n = n, sd = sd),
            y = rnorm(n = n, sd = sd),
            z = x * b_xz + y * b_yz + rnorm(n = n, sd = sd))
```

inaccurate estimate of X's effect on Y

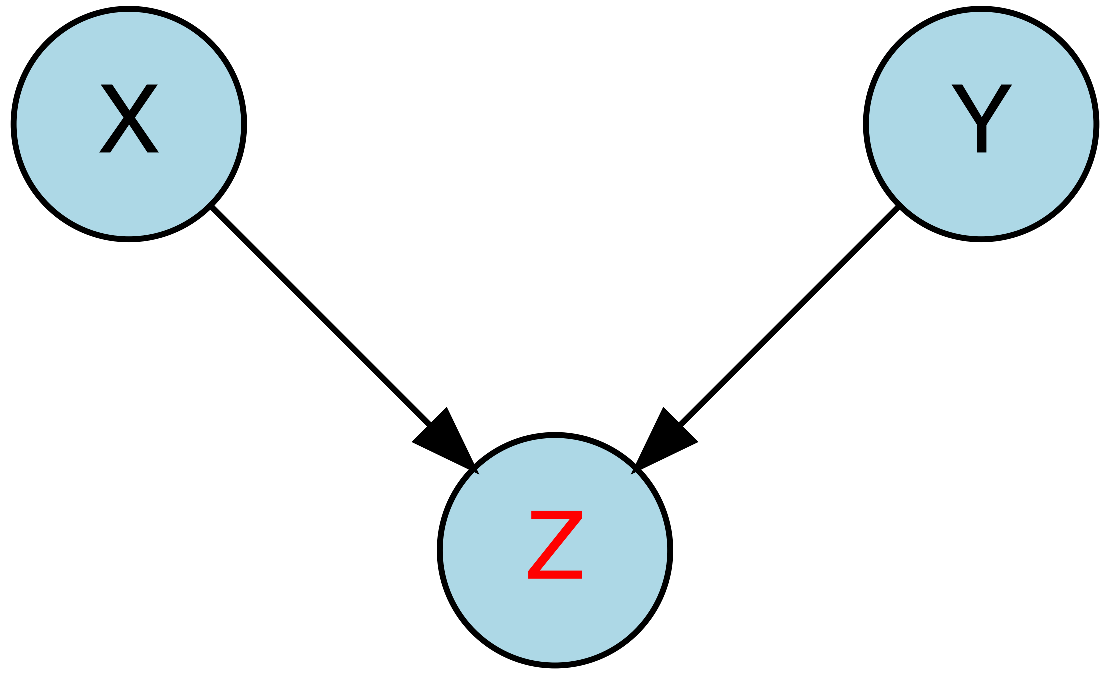
$$Y = b_0 + b_1 \cdot X + b_2 \cdot Z + e$$

```r
# with control
lm(formula = y ~ 1 + x + z,
   data = df) %>%
  summary()
```

```
Call:
lm(formula = y ~ x + z, data = df)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.35547 -0.30016  0.00298  0.31119  1.73408 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.009608   0.014477  -0.664    0.507    
x           -0.816164   0.018936 -43.102   <2e-16 ***
z            0.398921   0.006186  64.489   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4578 on 997 degrees of freedom
Multiple R-squared:  0.8066,	Adjusted R-squared:  0.8062 
F-statistic:  2079 on 2 and 997 DF,  p-value: < 2.2e-16
```

When should I control for variables?

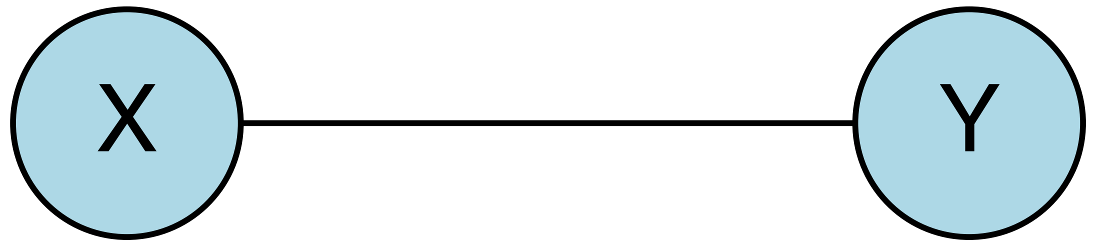
I want to estimate the effect that X has on Y



not controlling for Z

no path

controlling for Z



one path

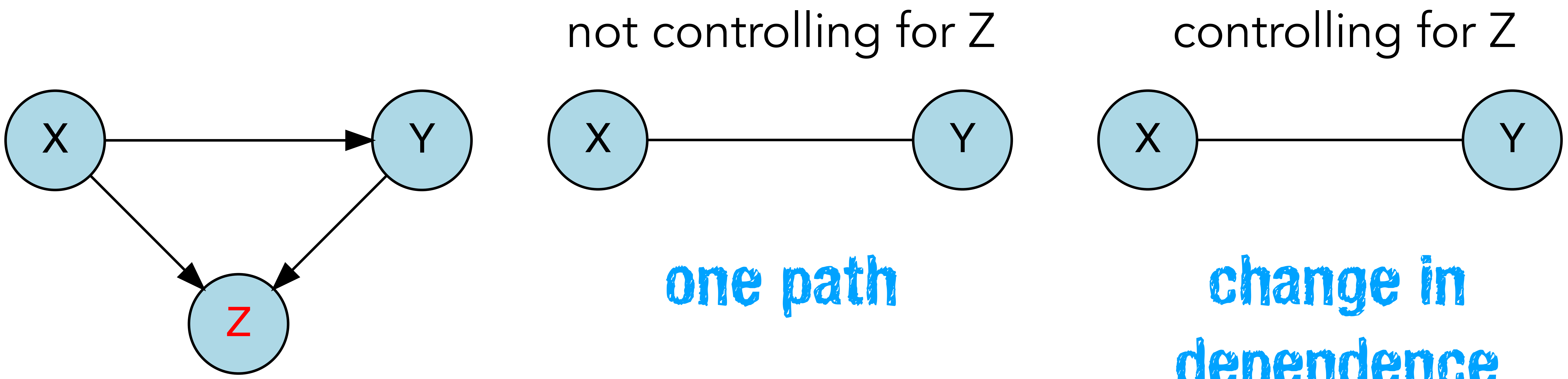
Z is a **bad** control here!

When should I control for variables?

- checking for **d-separation** tells us whether or not variables are (conditionally) independent
- it also tells us whether paths of dependence "open up", or get "closed down"
- the graphical procedure doesn't necessarily reveal whether the dependence between variables changes: it reveals the **structure** of dependence but not the **strength**
- you can always double check via running simulations in R

When should I control for variables?

I want to estimate the effect that X has on Y



one path

change in dependence

not revealed in the undirected graph

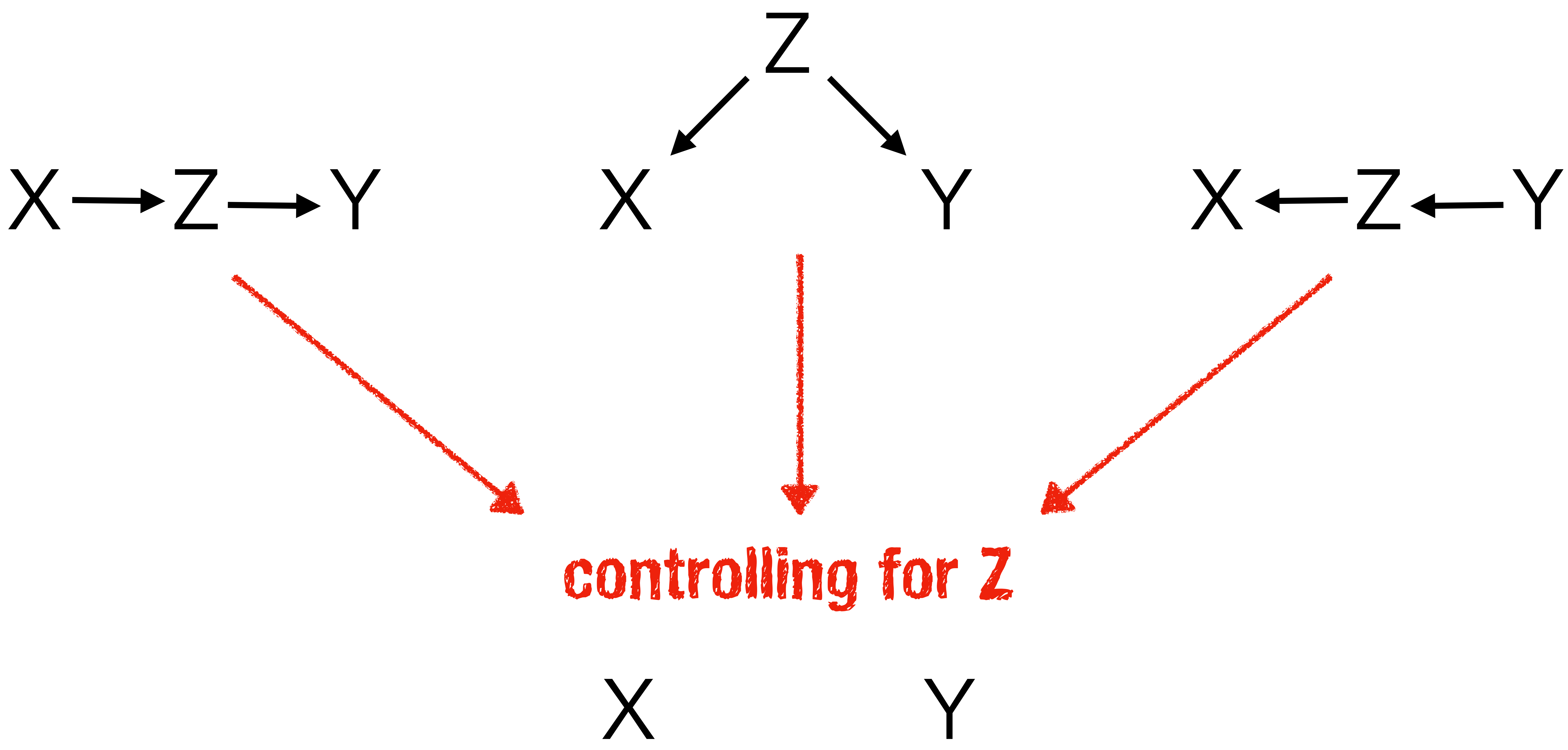
Z is a **bad** control here!

When should I control for variables?

- **good controls** reduce additional paths from `X` to `Y` apart from the direct path we are interested in estimating
- **bad controls** introduce additional paths (or change existing ones) that lead to a biased estimate of the direct path between `X` and `Y`

When should I control for variables?

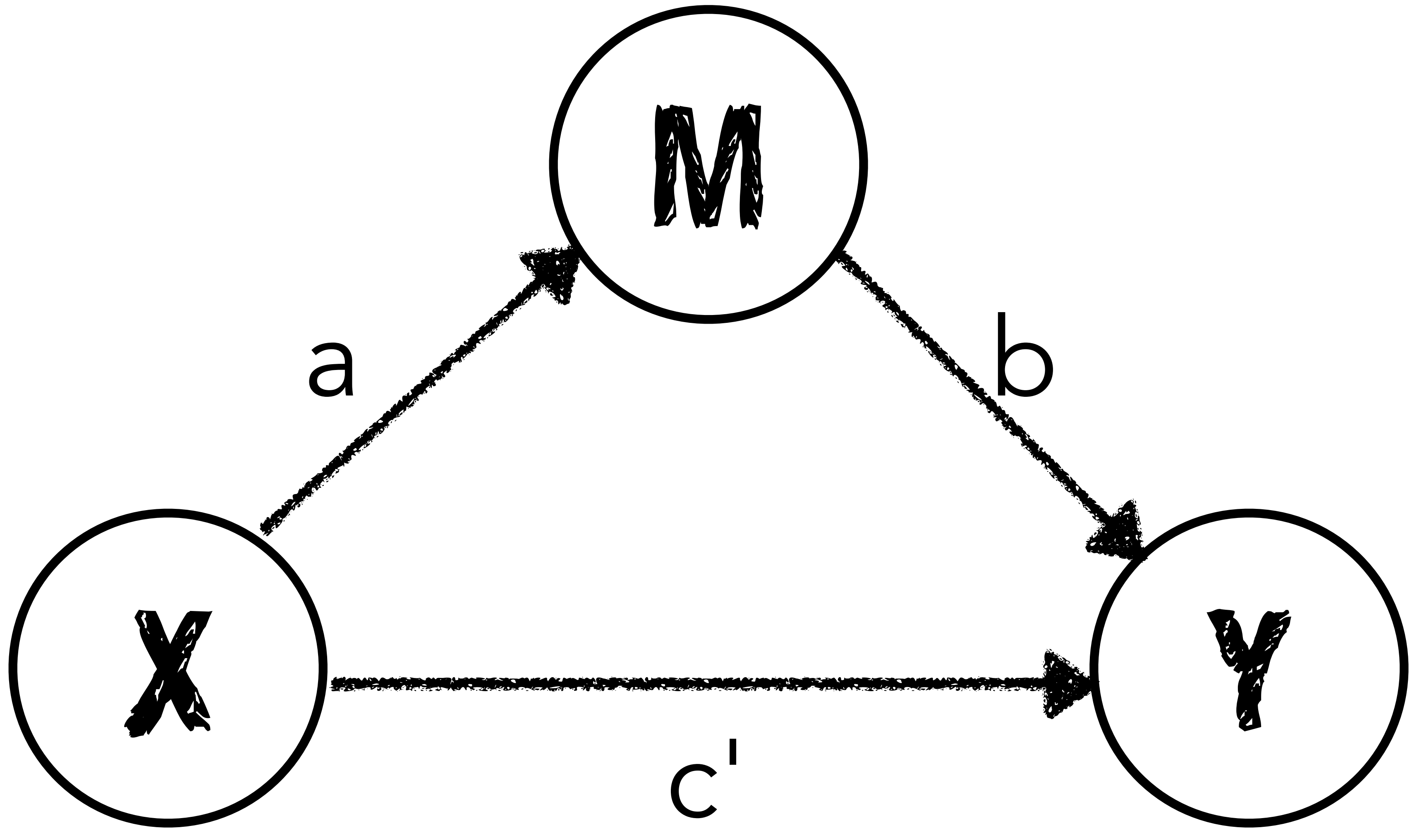
Problem: We don't know the ground truth ...



we need to manipulate X experimentally to tell these apart

Mediation

Definition

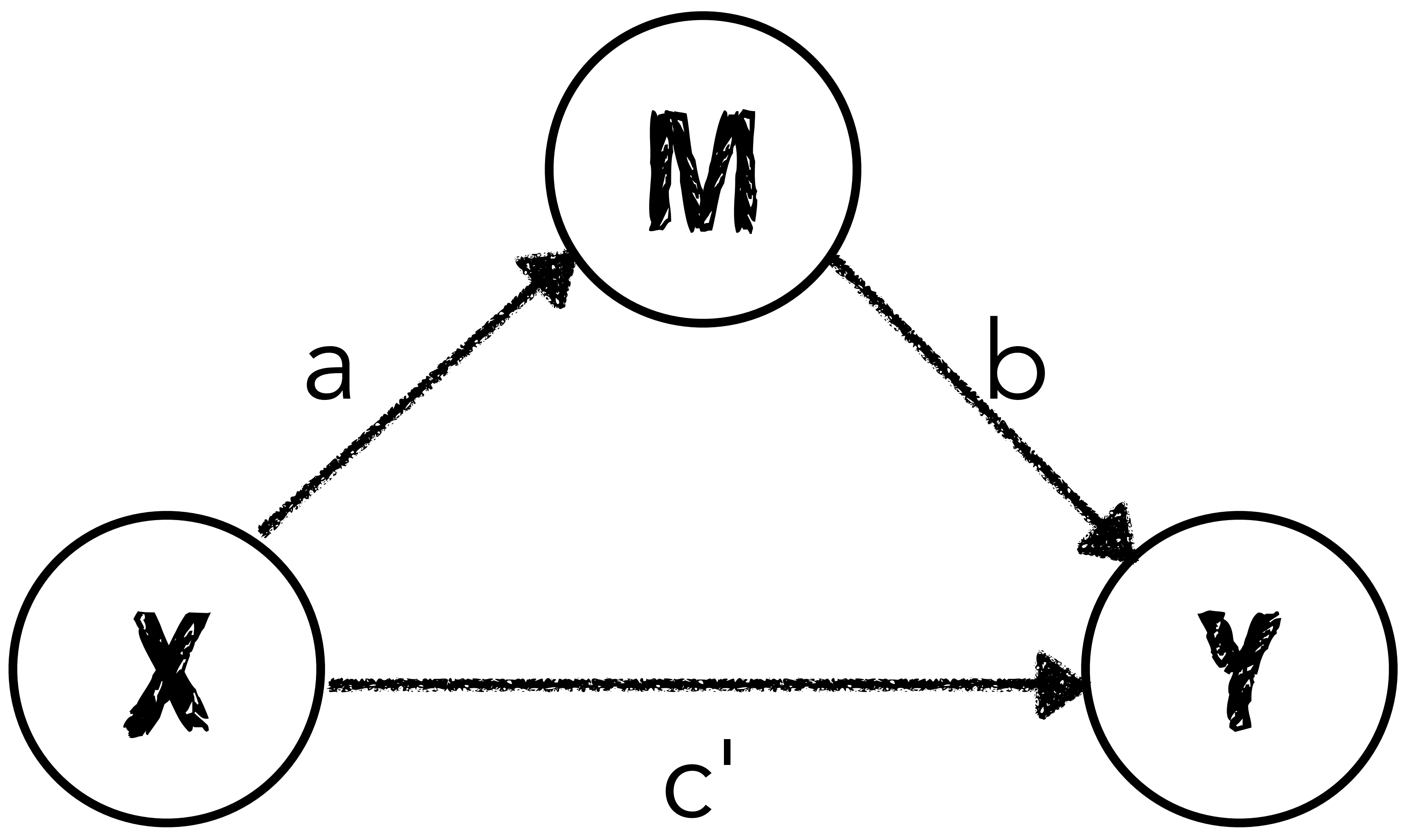


Rather than a direct causal relationship between **X** and **Y**, a mediation model proposes that **X** influences the mediator variable **M**, which in turn influences **Y**. Thus, the mediator variable serves to clarify the nature of the relationship between **X** and **Y**.

Adapted from Wikipedia

https://en.wikipedia.org/wiki/Mediation_(statistics)

Example



X = grades in Psych 252

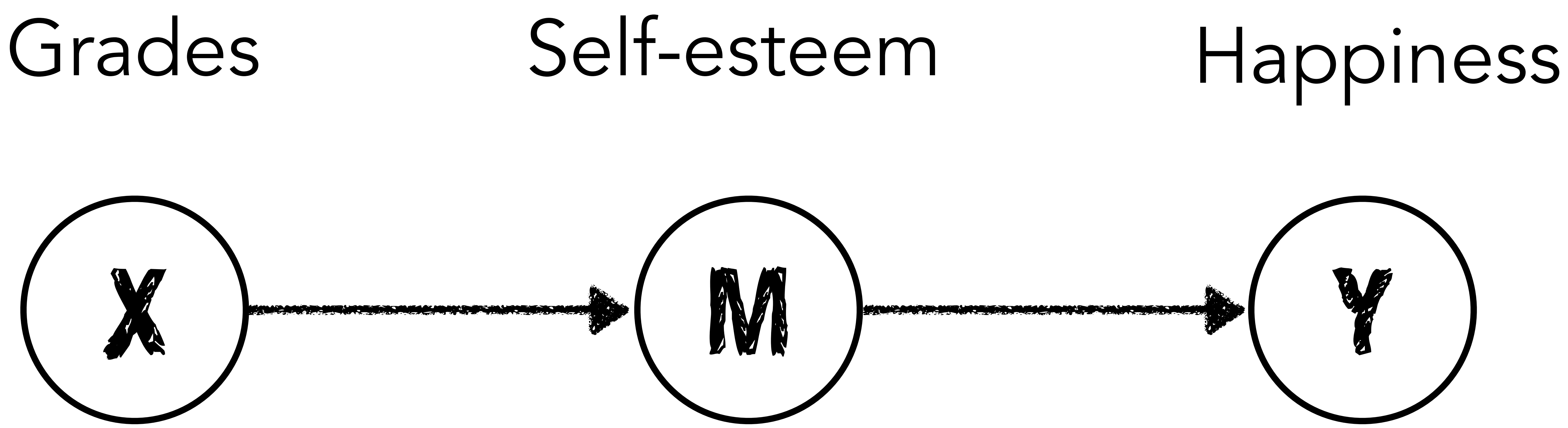
M = feelings of self-esteem

Y = happiness

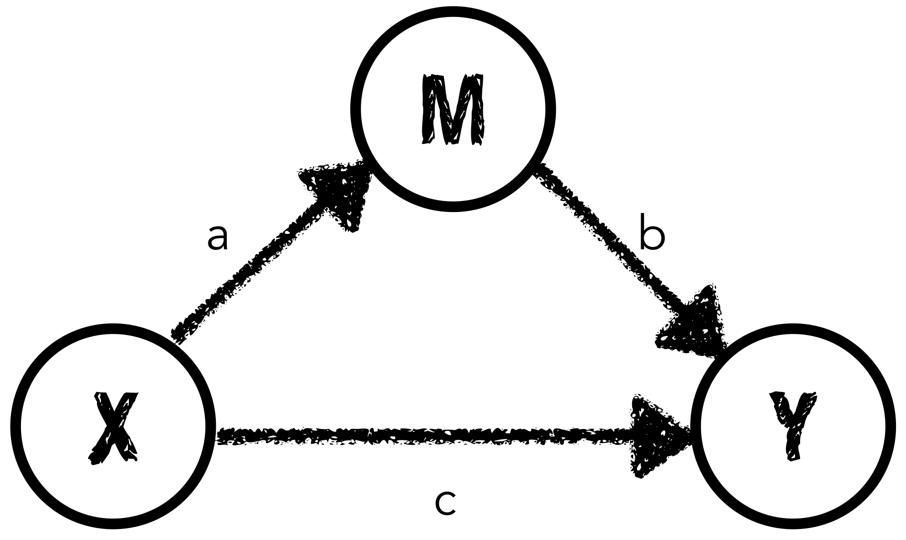
Is the relationship between grades in Psych 252 and happiness mediated by feelings of self-esteem?

Simulate a mediation analysis

```r
# number of participants
n = 100

# generate data
df.mediation = tibble(
  x = rnorm(n, 75, 7),              # grades
  m = 0.7 * x + rnorm(n, 0, 5), # self-esteem
  y = 0.4 * m + rnorm(n, 0, 5)  # happiness
)
```



Bootstrapping



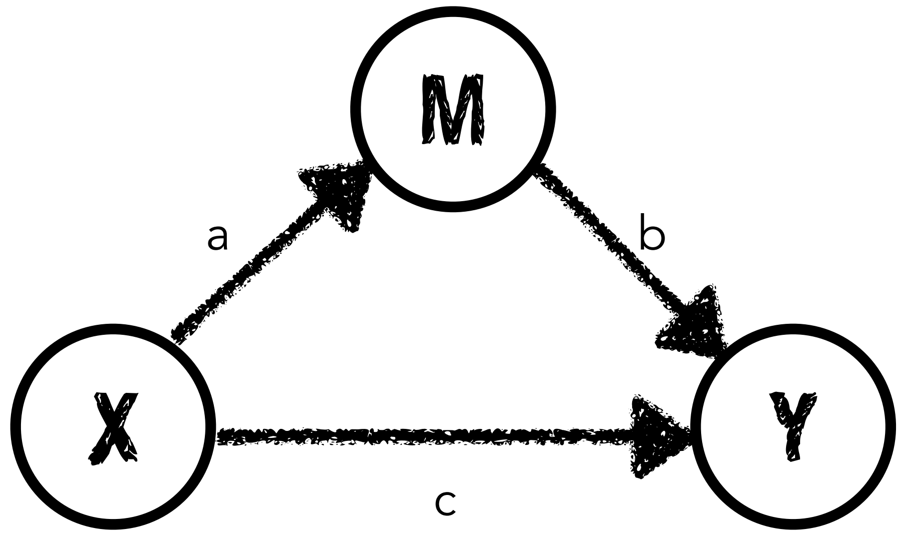

```r
library("mediation")

# bootstrapped mediation
fit.mediation = mediate(model.m = fit.m_x,
                        model.y = fit.y_mx,
                        treat = "x",
                        mediator = "m",
                        boot = T)

# summarize results
fit.mediation %>% summary()
```

$$\hat{m} = b_0 + b_1 \cdot x$$

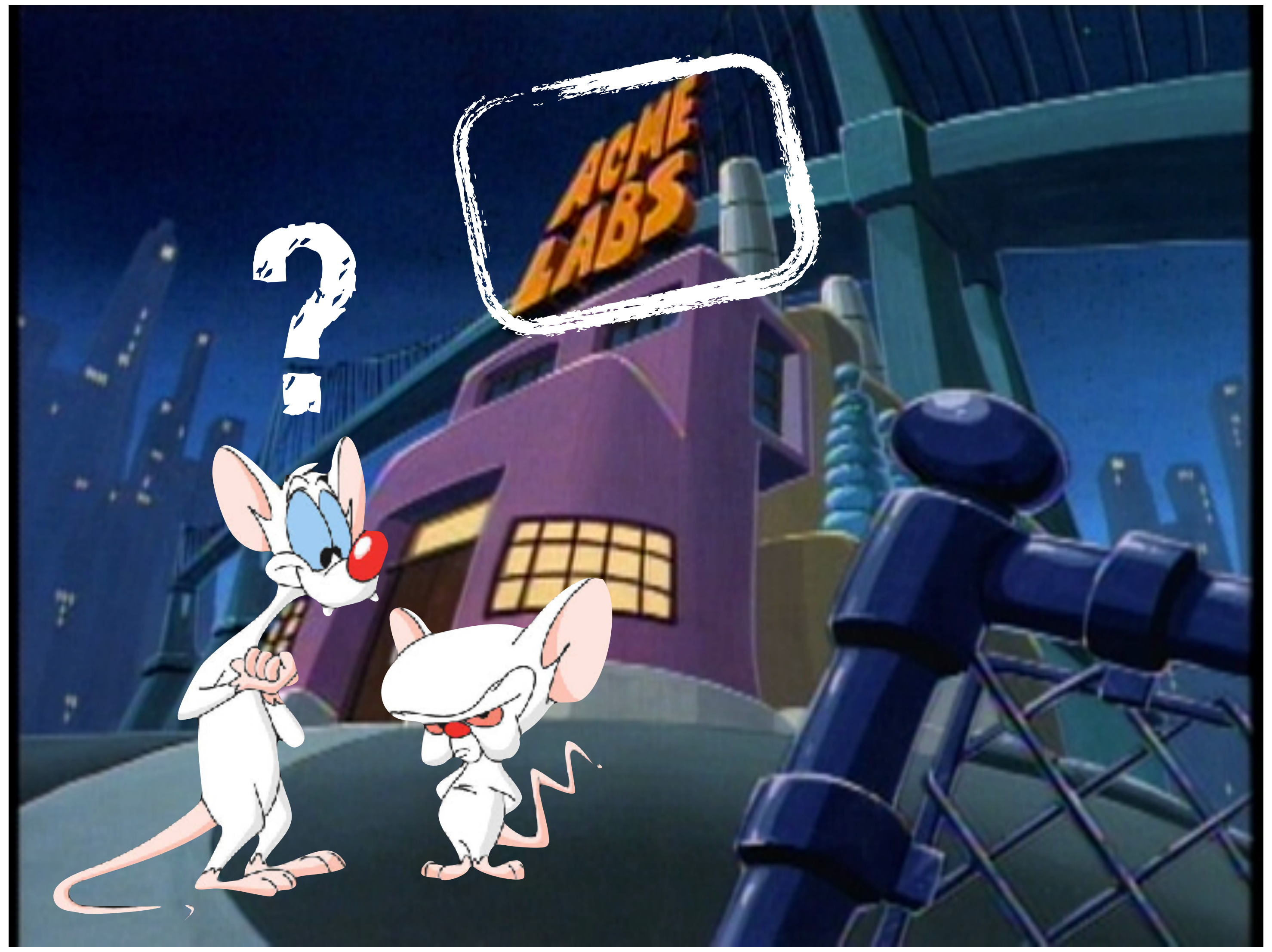
$$\hat{y} = b_0 + b_1 \cdot m + b_2 \cdot x$$

```
Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the Percentile Method

               Estimate 95% CI Lower 95% CI Upper p-value    
ACME            0.28078      0.14059         0.42  <2e-16 ***
ADE            -0.11179     -0.29276         0.10   0.272    
Total Effect    0.16899     -0.00415         0.34   0.064 .  
Prop. Mediated  1.66151     -3.22476        11.46   0.064 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 100


Simulations: 1000
```

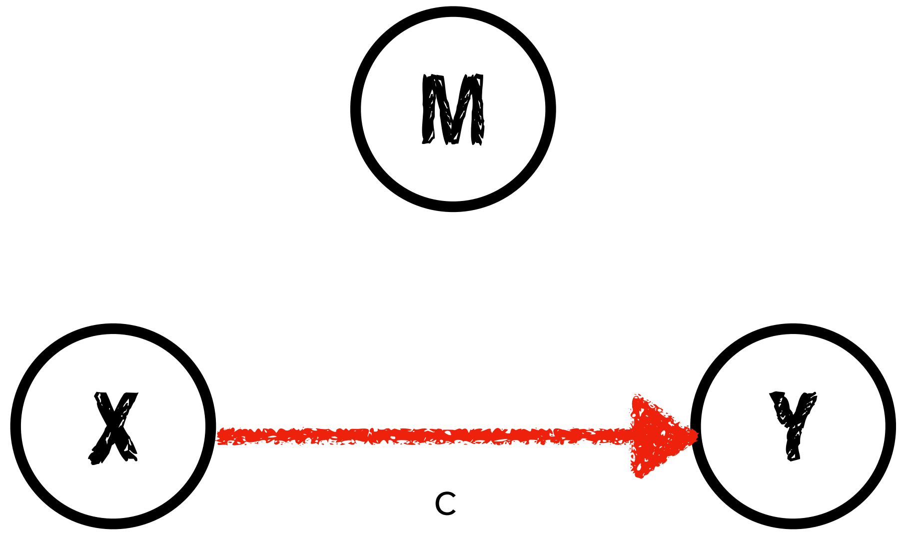
2. Bootstrapping




```
Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the Percentile Method

               Estimate 95% CI Lower 95% CI Upper p-value
ACME            0.28078      0.14059         0.42  <2e-16 ***
ADE            -0.11179     -0.29276         0.10   0.272
Total Effect    0.16899     -0.00415         0.34   0.064 .
Prop. Mediated  1.66151     -3.22476        11.46   0.064 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 100


Simulations: 1000
```



2. Bootstrapping

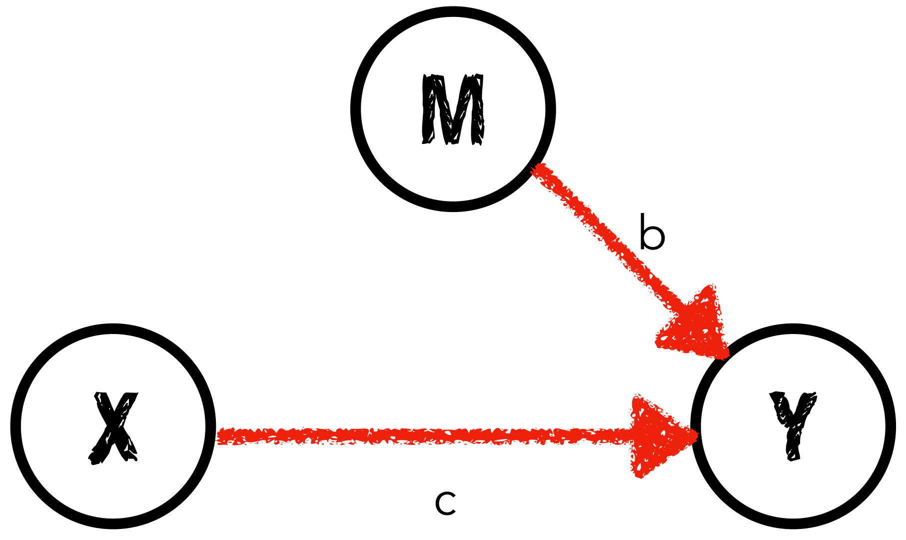



```
Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the Percentile Method

               Estimate 95% CI Lower 95% CI Upper p-value    
ACME            0.28078      0.14059         0.42  <2e-16 ***
ADE            -0.11179     -0.29276         0.10   0.272    
Total Effect    0.16899     -0.00415         0.34   0.064 .  
Prop. Mediated  1.66151     -3.22476        11.46   0.064 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 100 


Simulations: 1000 
```

$$\hat{y} = b_0 + b_1 \cdot x$$

```
Call:
lm(formula = y ~ 1 + x, data = df.mediation)

Residuals:
    Min      1Q  Median      3Q     Max 
-10.917  -3.738  -0.259   2.910  12.540 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  8.78300    6.16002   1.426   0.1571  
x            0.16899    0.08116   2.082   0.0399 *
```

2. Bootstrapping




```
Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the Percentile Method

               Estimate 95% CI Lower 95% CI Upper p-value    
ACME            0.28078      0.14059         0.42  <2e-16 ***
ADE            -0.11179     -0.29276         0.10   0.272    
Total Effect    0.16899     -0.00415         0.34   0.064 .  
Prop. Mediated  1.66151     -3.22476        11.46   0.064 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 100


Simulations: 1000
```

$$\hat{y} = b_0 + b_1 \cdot m + b_2 \cdot x$$ **ADE:** Average direct effect

```
Call:
lm(formula = y ~ 1 + m + x, data = df.mediation)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.3651 -3.3037 -0.6222  3.1068 10.3991 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.80952    5.68297   1.374    0.173    
m            0.42381    0.09899   4.281 4.37e-05 ***
x           -0.11179    0.09949  -1.124    0.264    
```

2. Bootstrapping

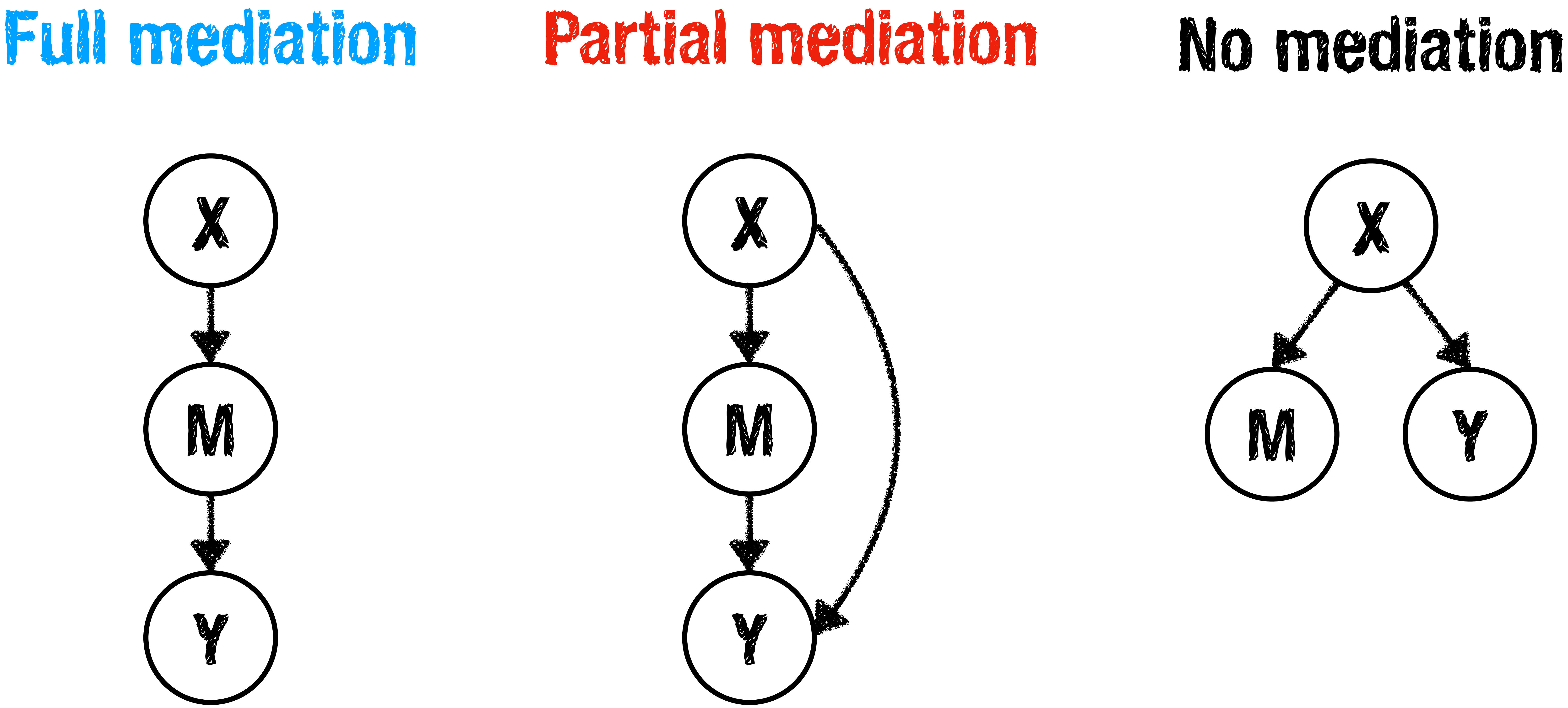
```
Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the Percentile Method

               Estimate 95% CI Lower 95% CI Upper p-value    
ACME            0.28078      0.14059         0.42  <2e-16 ***
ADE            -0.11179     -0.29276         0.10   0.272    
Total Effect    0.16899     -0.00415         0.34   0.064 .  
Prop. Mediated  1.66151     -3.22476        11.46   0.064 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 100


Simulations: 1000
```

ACME: Average causal mediation effect

ACME = Total effect - ADE

ADE: Average direct effect

Underlying causal model

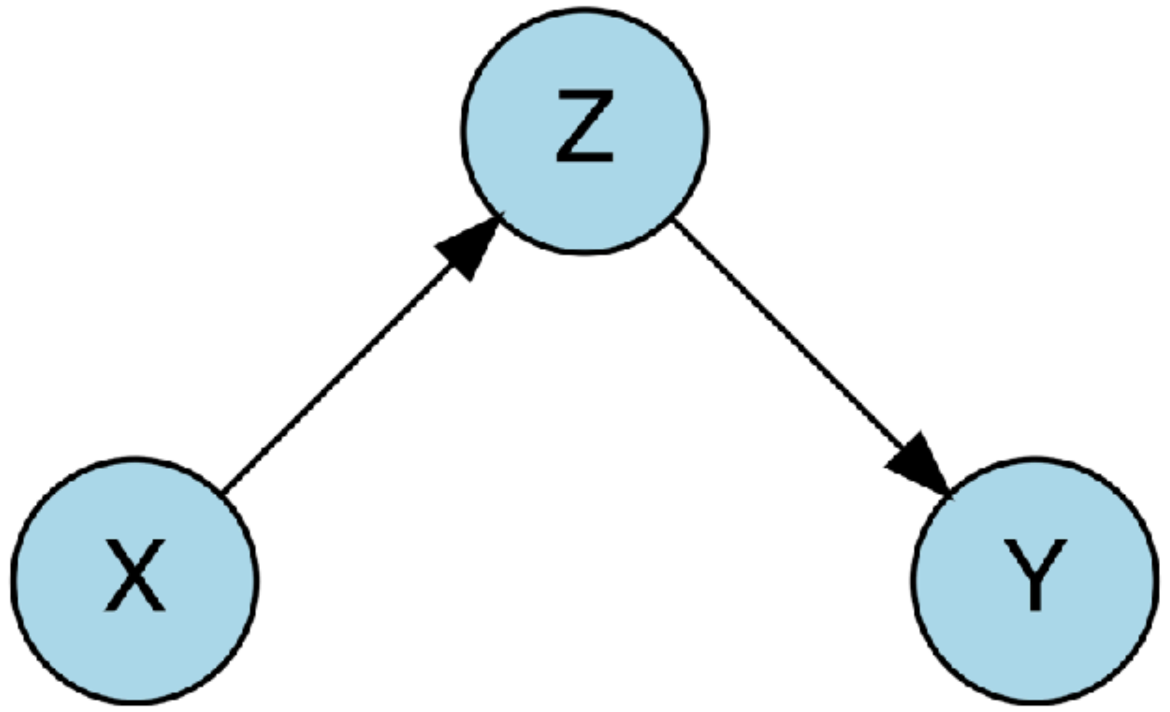


Full mediation: When the effect of **X** on **Y** completely disappears, **M** fully mediates between **X** and **Y**.

Partial mediation: When the effect of **X** on **Y** still exists, but in a smaller magnitude, **M** partially mediates between **X** and **Y**.

Beware of mediation analyses!!!

```r
set.seed(1)

n = 100 # number of observations

# causal chain
df.causal_chain = tibble(x = rnorm(n, 0, 1),
                         z = 2 * x + rnorm(n, 0, 1),
                         y = 2 * z + rnorm(n, 0, 1))
```

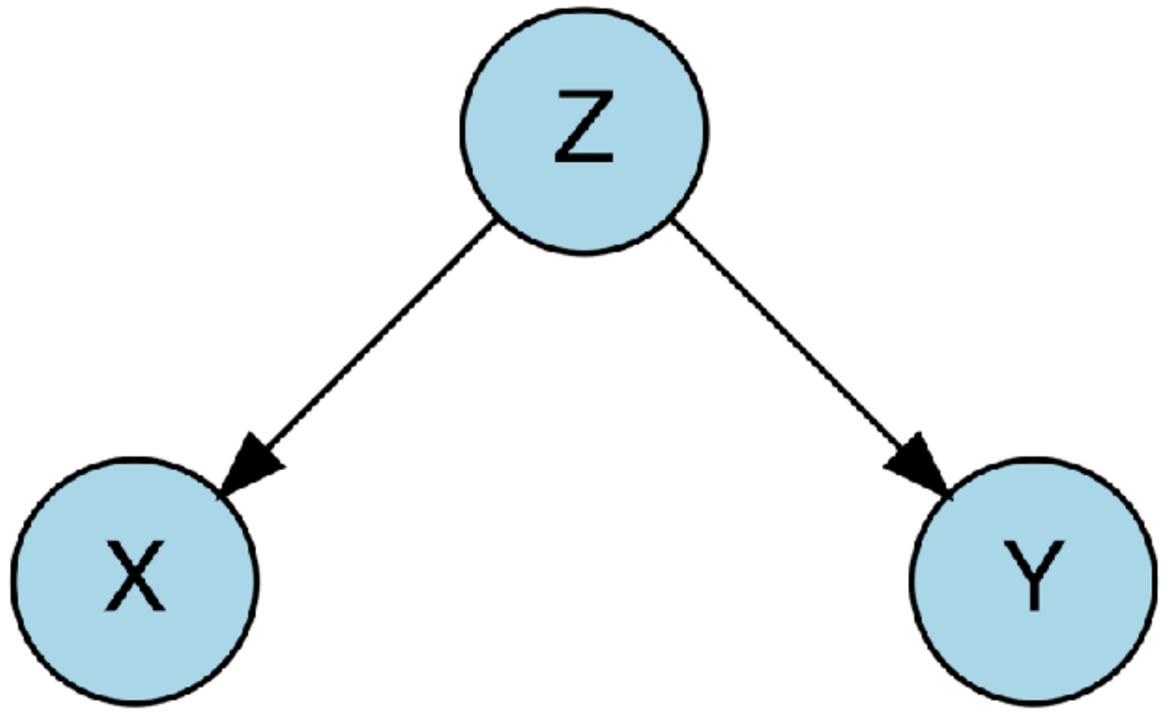



```
Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the Percentile Method

               Estimate 95% CI Lower 95% CI Upper p-value
ACME             0.8287       0.6234         1.05  <2e-16 ***
ADE             -0.0535      -0.2548         0.15    0.55
Total Effect     0.7752       0.6391         0.90  <2e-16 ***
Prop. Mediated   1.0690       0.8131         1.35  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 100


Simulations: 1000
```

nice mediation result!

Beware of mediation analyses!!!

```r
set.seed(1)

n = 100 # number of observations

# common cause
df.common_cause = tibble(z = rnorm(n, 0, 1),
                         x = 2 * z + rnorm(n, 0, 1),
                         y = 2 * z + rnorm(n, 0, 1))
```




```
Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the Percentile Method

               Estimate 95% CI Lower 95% CI Upper p-value    
ACME             0.8287       0.6065         1.04  <2e-16 ***
ADE             -0.0535      -0.2675         0.16    0.56    
Total Effect     0.7752       0.6353         0.90  <2e-16 ***
Prop. Mediated   1.0690       0.8134         1.37  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sample Size Used: 100


Simulations: 1000
```

(not) nice mediation result!

Limitations

- correlational analysis
  - we need theories / experiments to tease apart causes and effects to properly map our variables onto the diagram

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

ELSEVIER

Reports

What mediation analysis can (not) do ☆

Klaus Fiedler [a,*], Malte Schott [a], Thorsten Meiser [b]

[a] University of Heidelberg, Germany
[b] University of Mannheim, Germany

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, 47(6), 1231-1236.

Limitations

many-to-one mapping

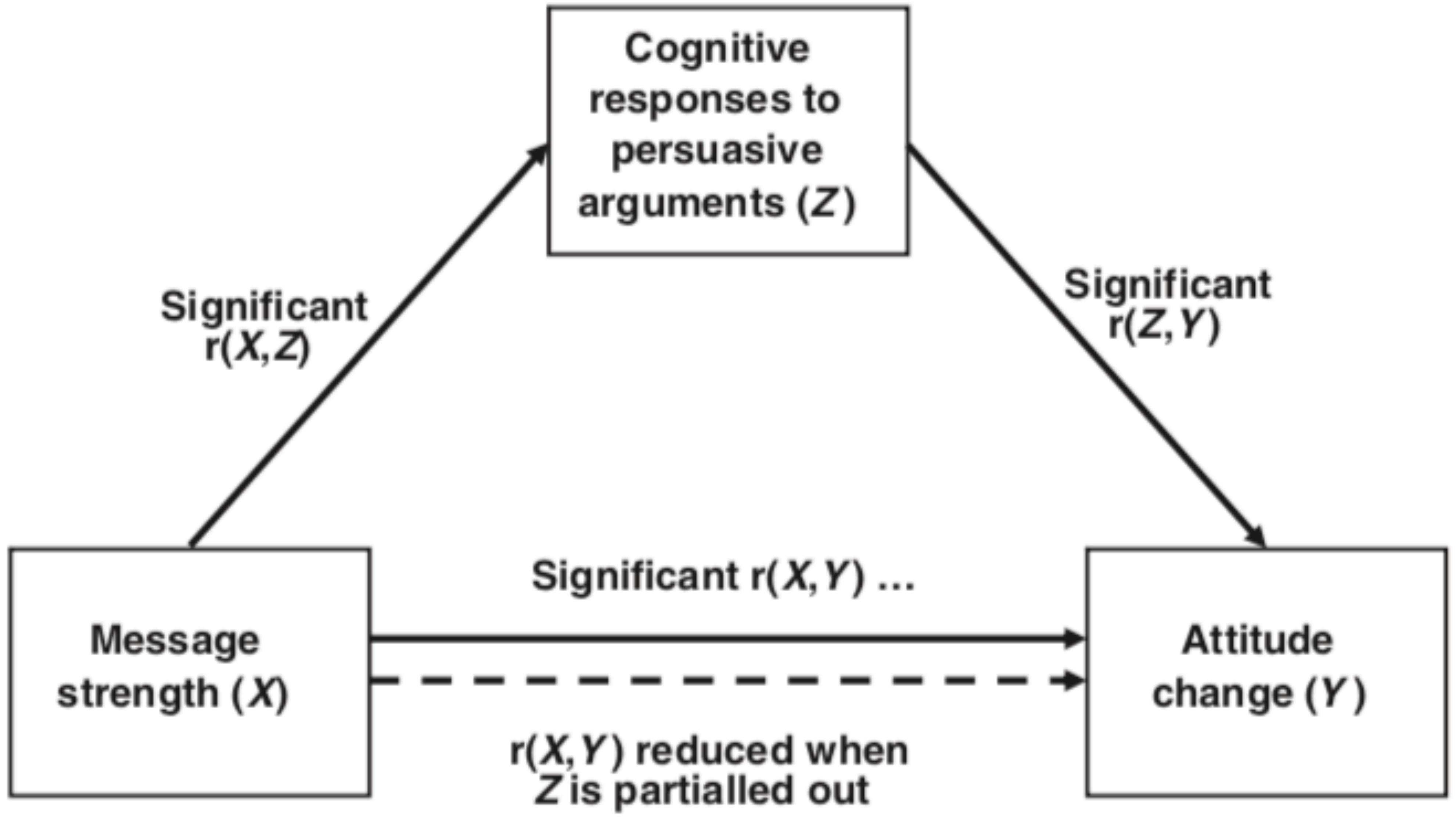


Fig. 1. Illustration of a mediation analysis in the context of persuasion research: The impact of message strength (X) on attitude change (Y) is supposed to be mediated by the recipients' cognitive responses (Z).

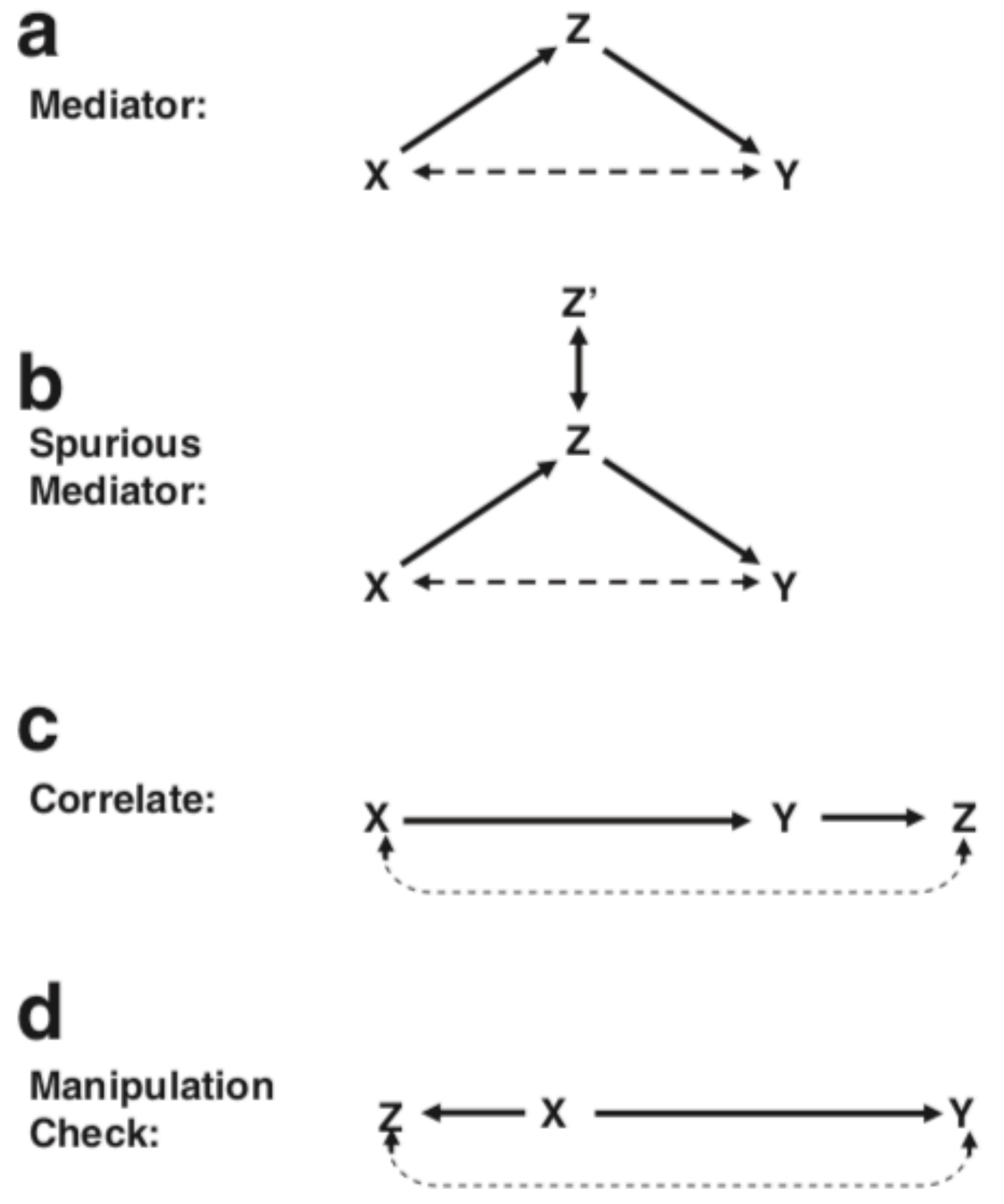


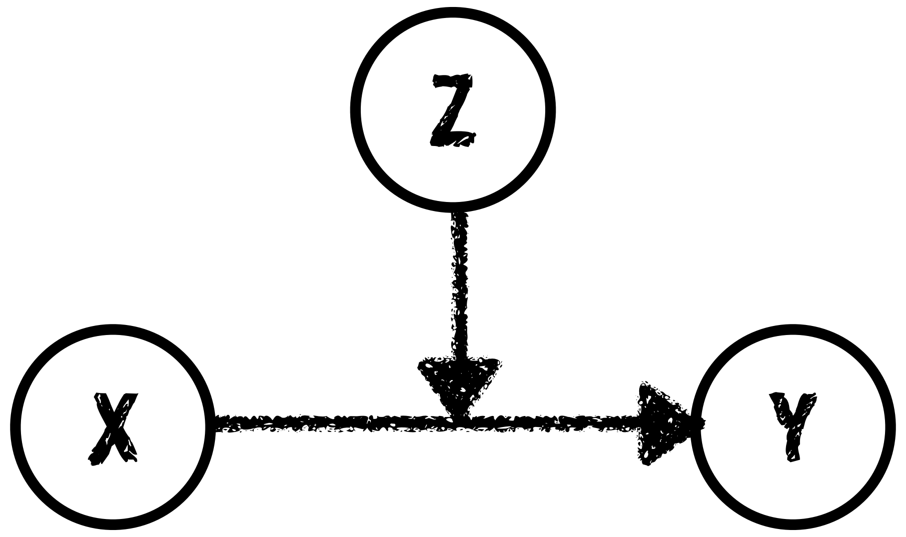
Fig. 2. Different causal models for the interpretation of the same intercorrelations between three variables, X, Y, and Z.

only experiments allow us to tell apart possible causal structures

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, 47(6), 1231-1236.

Moderation

Definition

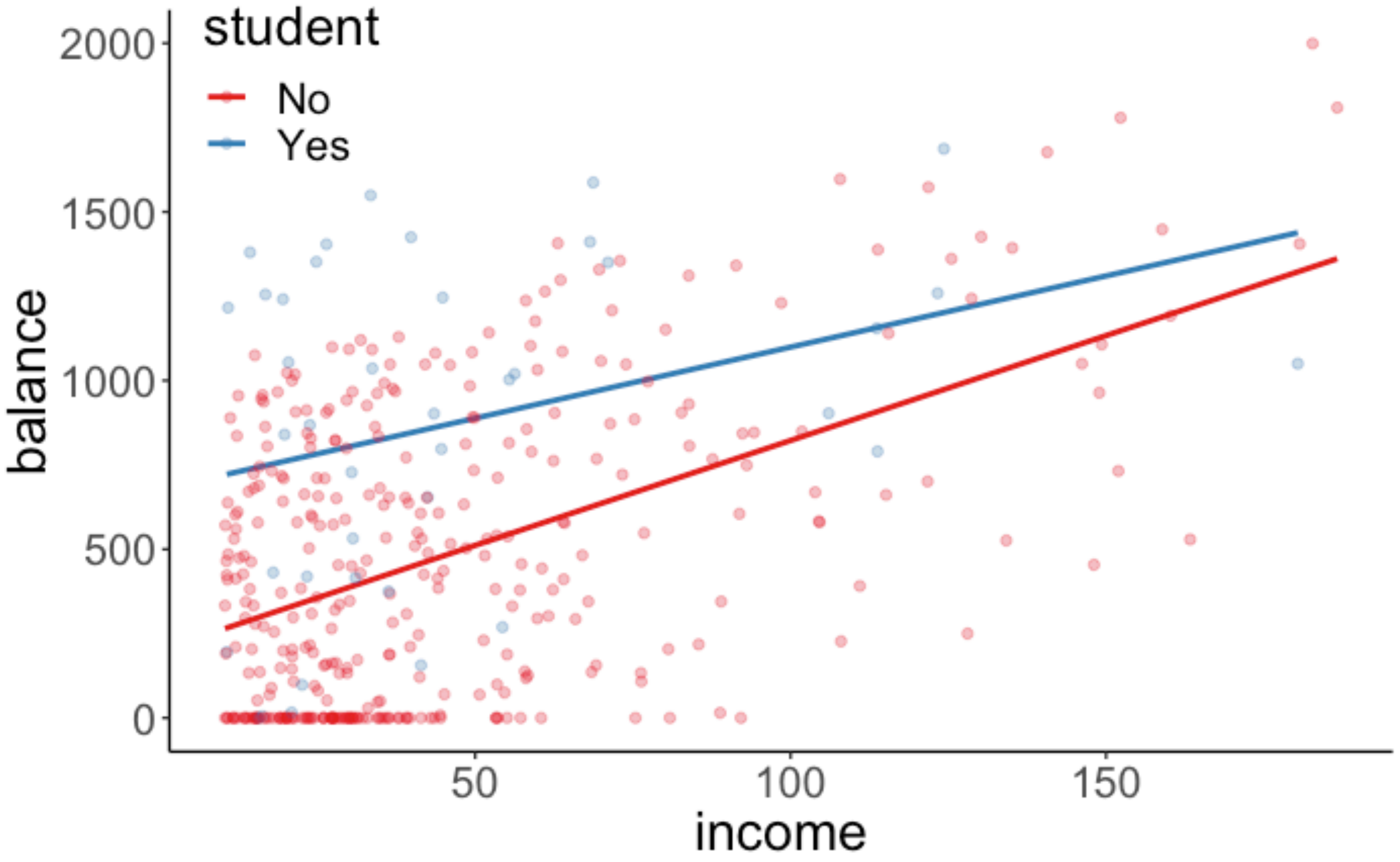


Moderation means that the effect of a predictor depends on the value of another.

Here, the nature of the relationship between **X** and **Y** depends on **Z**.

Have we come across moderation already?

Relationship between credit card balance, income, and whether the person is a student.



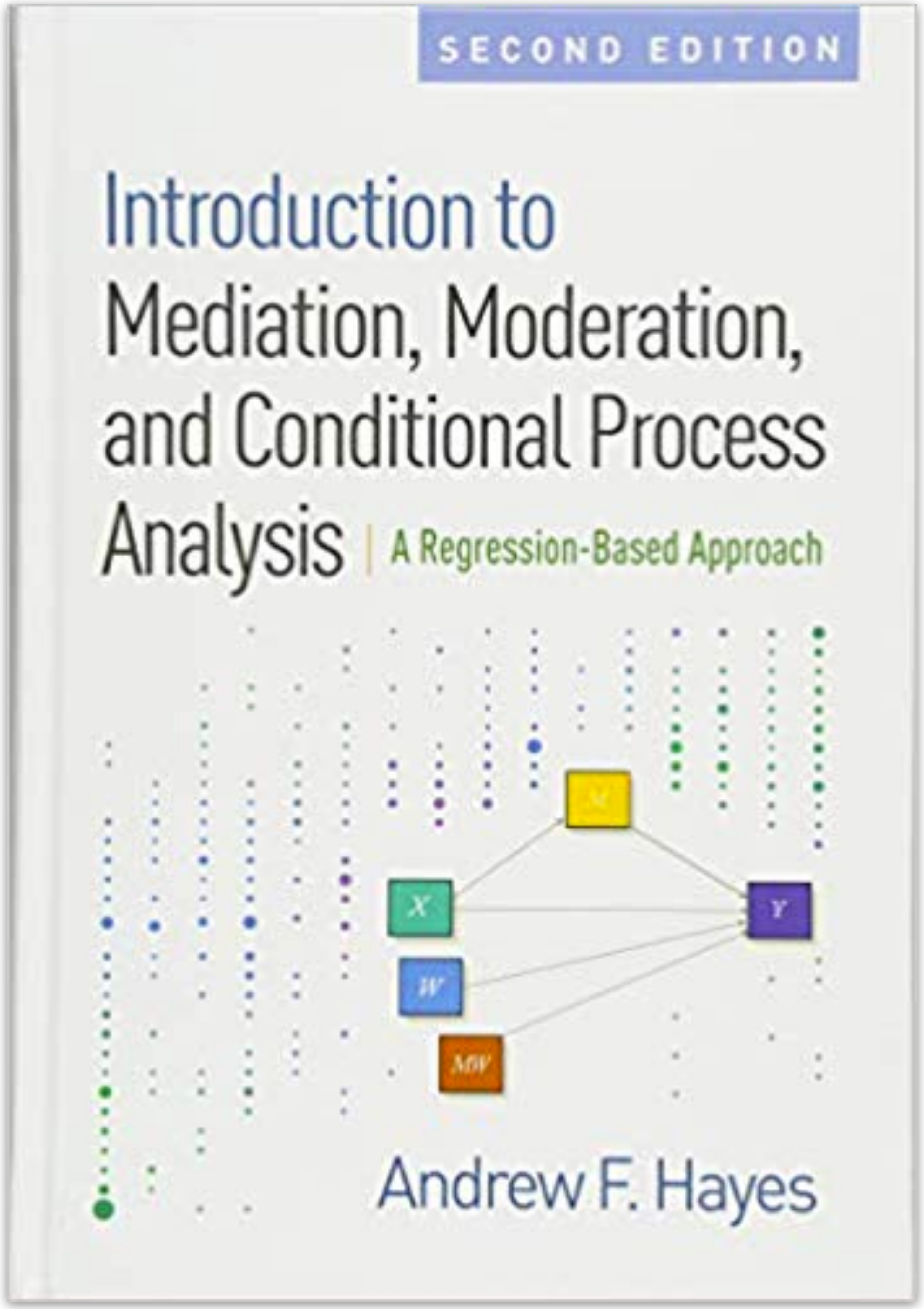
$$\widehat{\text{balance}}_i = 200.62 + 6.22 \cdot \text{income}_i + 476.68 \cdot \text{student}_i - 2.00 \cdot (\text{income}_i \times \text{student}_i)$$

if student = "No" $\quad \widehat{\text{balance}}_i = 200.62 + 6.22 \cdot \text{income}_i$

if student = "Yes"

$$\begin{aligned}\widehat{\text{balance}}_i &= 200.62 + 6.22 \cdot \text{income}_i + 476.68 \cdot 1 - 2.00 \cdot (\text{income}_i \times 1)\\ &= 677.3 + 6.22 \cdot \text{income}_i - 2.00 \cdot \text{income}_i\\ &= 677.3 + 4.22 \cdot \text{income}_i\end{aligned}$$

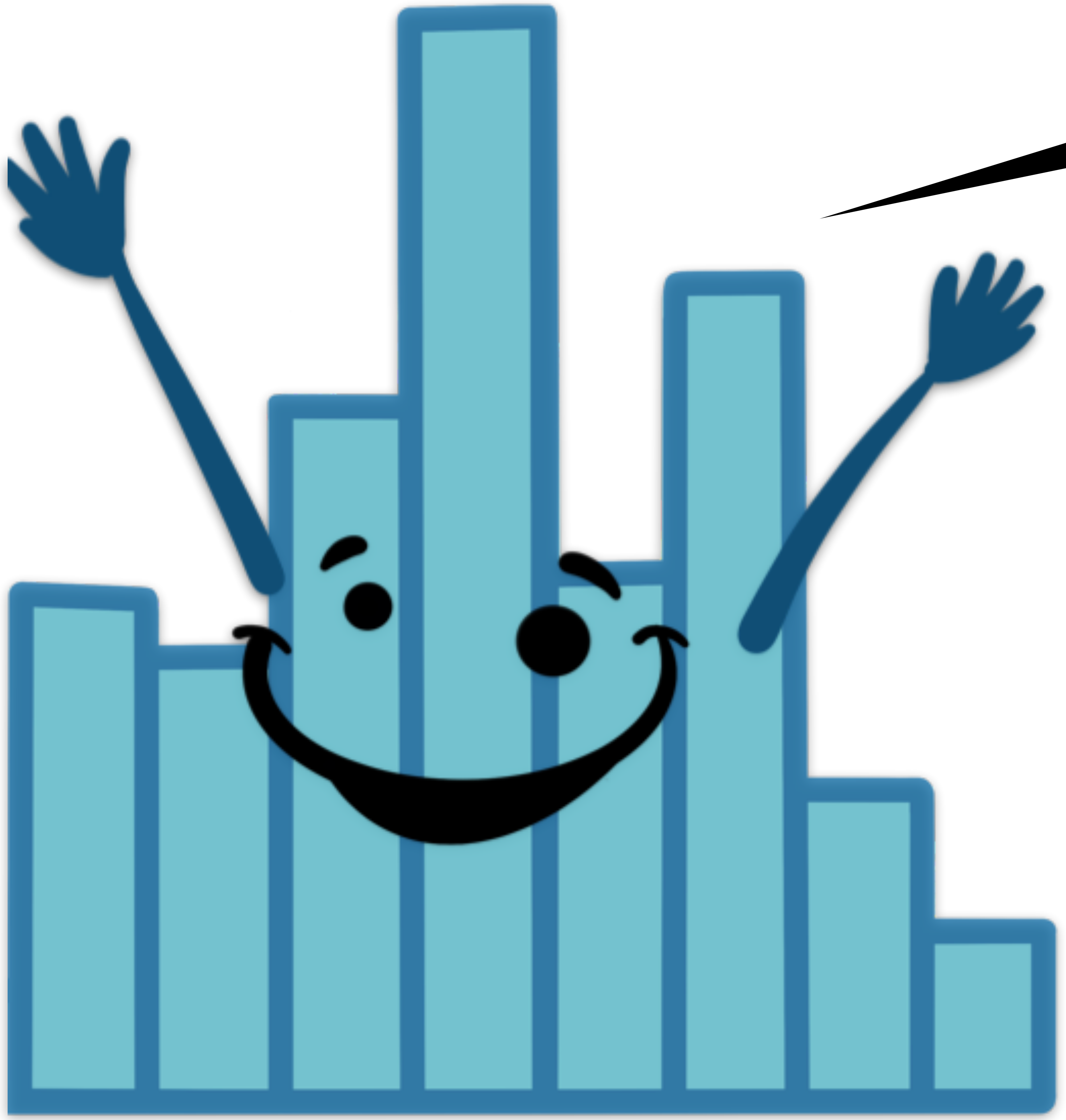
Learn more about mediation and moderation



Recoded with `brms` by Solomon Kurz here: https://bookdown.org/connect/#/apps/1523/access

02:00

stretch break!



Bayesian Data Analysis

Datacamp course

recommended!!

https://www.datacamp.com/courses/fundamentals-of-bayesian-data-analysis-in-r

Great online book

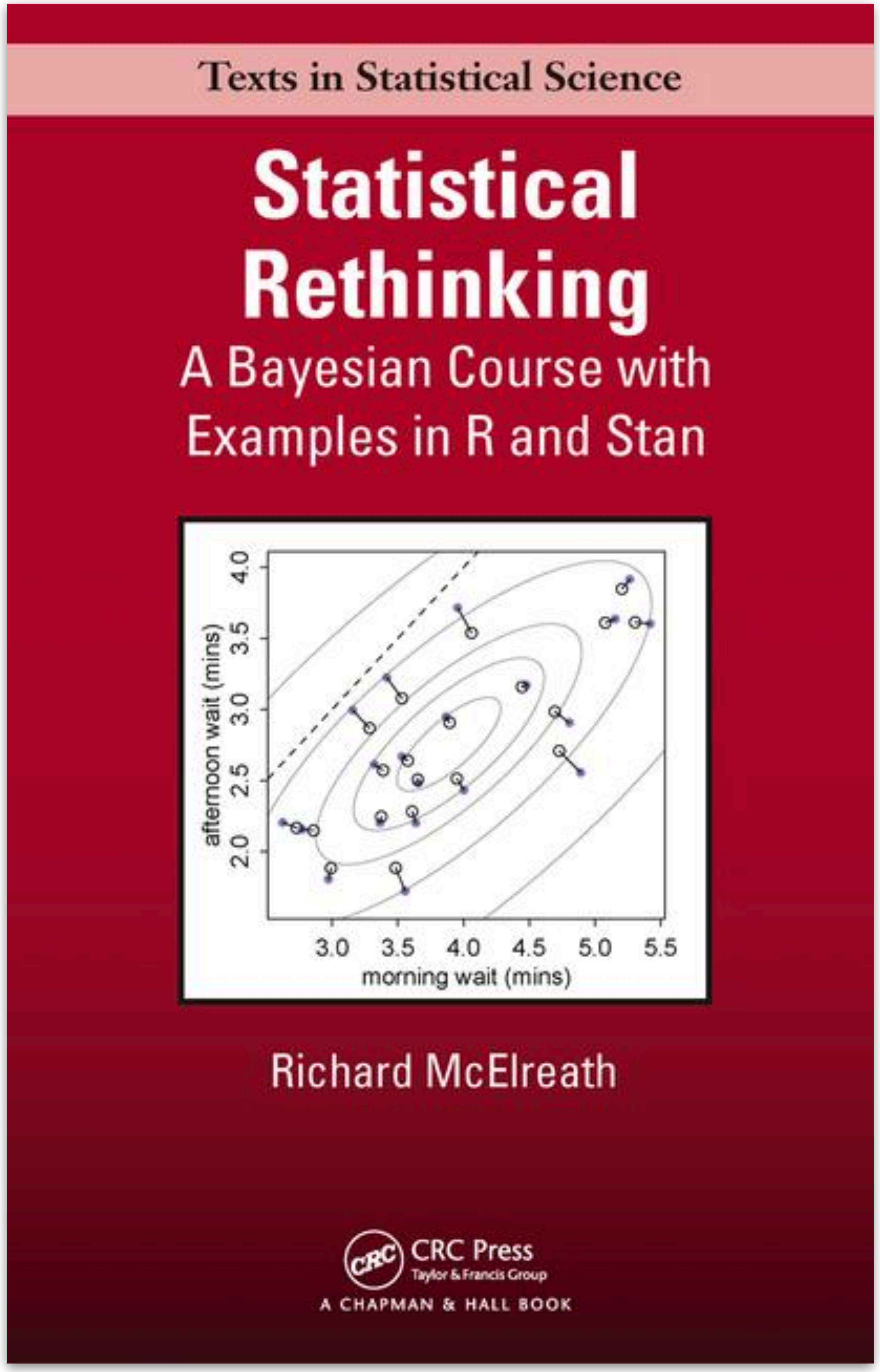
An Introduction to Data Analysis

Michael Franke

last rendered at: 2023-02-08 14:03:16

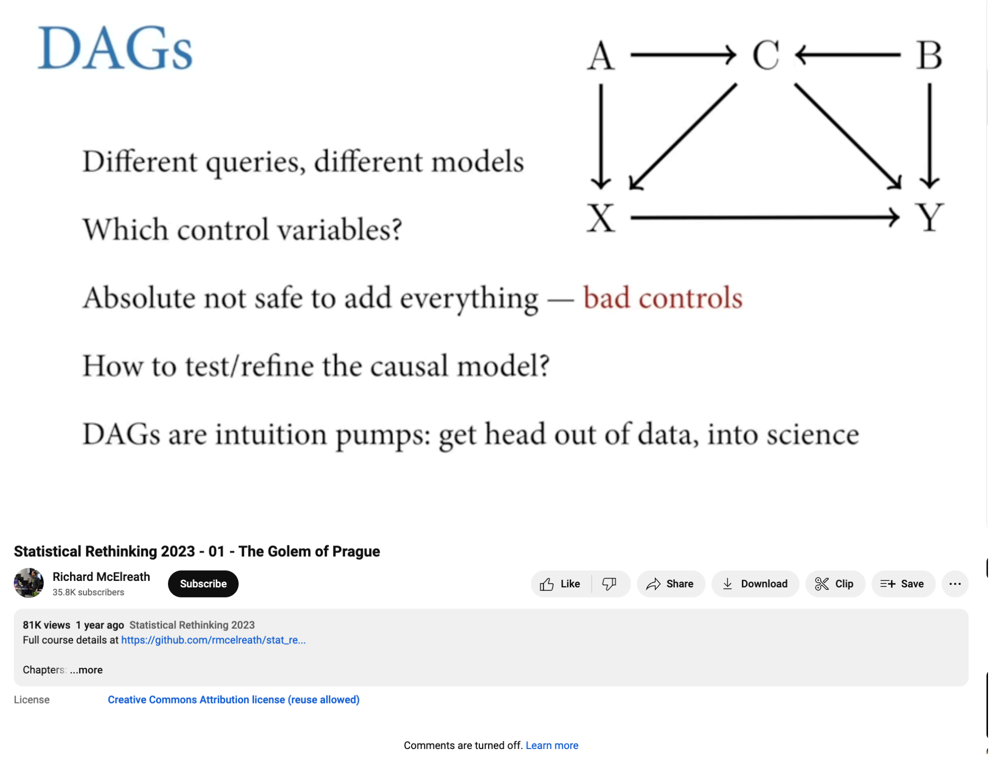
https://michael-franke.github.io/intro-data-analysis/index.html

Great book on Bayesian data analysis



- nice hands-on book (which uses R throughout)
- rewrite of all the code with "`tidyverse`" and "`BRMS`" is here: https://bookdown.org/content/4857/
- video lectures are available here: https://github.com/rmcelreath/stat_rethinking_2025

video lectures available



Comparison between frequentist and Bayesian data analysis

Goal of data analysis: Inference about the world

Frequentist statistics

- generate a sampling distribution of the test statistic assuming H_0
- compare observed value of the test statistic with the sampling distribution
- reject the H_0 if probability of observed value (or more extreme values) is less than α

Bayesian statistics

- directly test hypotheses of interest
- define prior over hypotheses **p(H)**
- compute likelihood of the data for each hypothesis **p(D|H)**
- use Bayes' rule to infer the posterior over hypotheses given the data **p(H|D)**

Objections to frequentist NHST

null hypothesis significance testing

- p-value is not a measure of evidential support
  - becomes smaller as *N* increases
- results are often misinterpreted (both p-values and confidence intervals are not particularly intuitive)
- what we want to know: $p(\text{Hypothesis} \mid \text{Data})$
- what we calculate: $p(\text{Data} \mid \text{Null Hypothesis})$

Wasserstein, Schirm, & Lazar (2019) Moving to a World Beyond p < 0.05. *The American Statistician*

Frequentists vs. Bayesian

- both "want" to evaluate the evidence for a hypothesis using a sample of data **p(H|D)**
- it's often easier to calculate the inverse: the probability of the data given a hypothesis **p(D|H)**
- frequentists use a rule of thumb (p-value) to make a decision
- Bayesians use Bayes' rule

Why don't more people use Bayesian Statistics?

- supposedly more difficult
  - relies on the logic of probability theory
- reliance on a *prior*
- reliance on computing and simulation
  - we can't just use SPSS
  - but we can use JASP (Just Another Statistics Program)

and we've already learned how to simulate and visualize data in this class!

What are (some of) the benefits of Bayesian data analysis?

- intuitive model testing and comparison
- straightforward interpretation of results
  - Bayesian credible intervals vs. Confidence intervals
- more model flexibility
  - adequately express assumptions about the data-generating process
- better predictions!

Flash from the past

Clue guide to probability

we derived this using the definition of conditional probability

$$p(B \mid A) = \frac{p(A \mid B) \cdot p(B)}{p(A)}$$

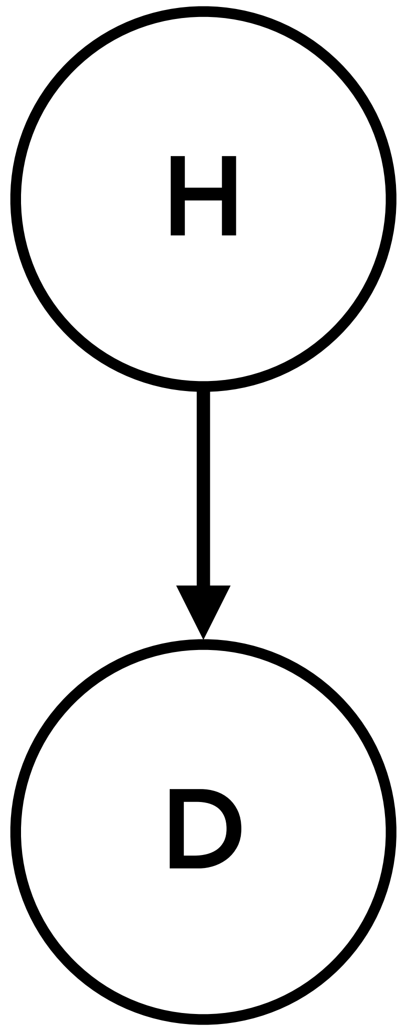
posterior, likelihood, prior

$$p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{p(D)}$$

subjective probability interpretation

H = Hypothesis

D = Data



formal framework for learning from data

updating our prior belief $p(H)$, to a posterior belief $p(H \mid D)$ given some data

Clue guide to probability

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is "95% accurate"; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that **$P(T|D) = 0.95$** and **$P(\neg T|\neg D) = 0.95$**. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

Clue guide to probability

what we know

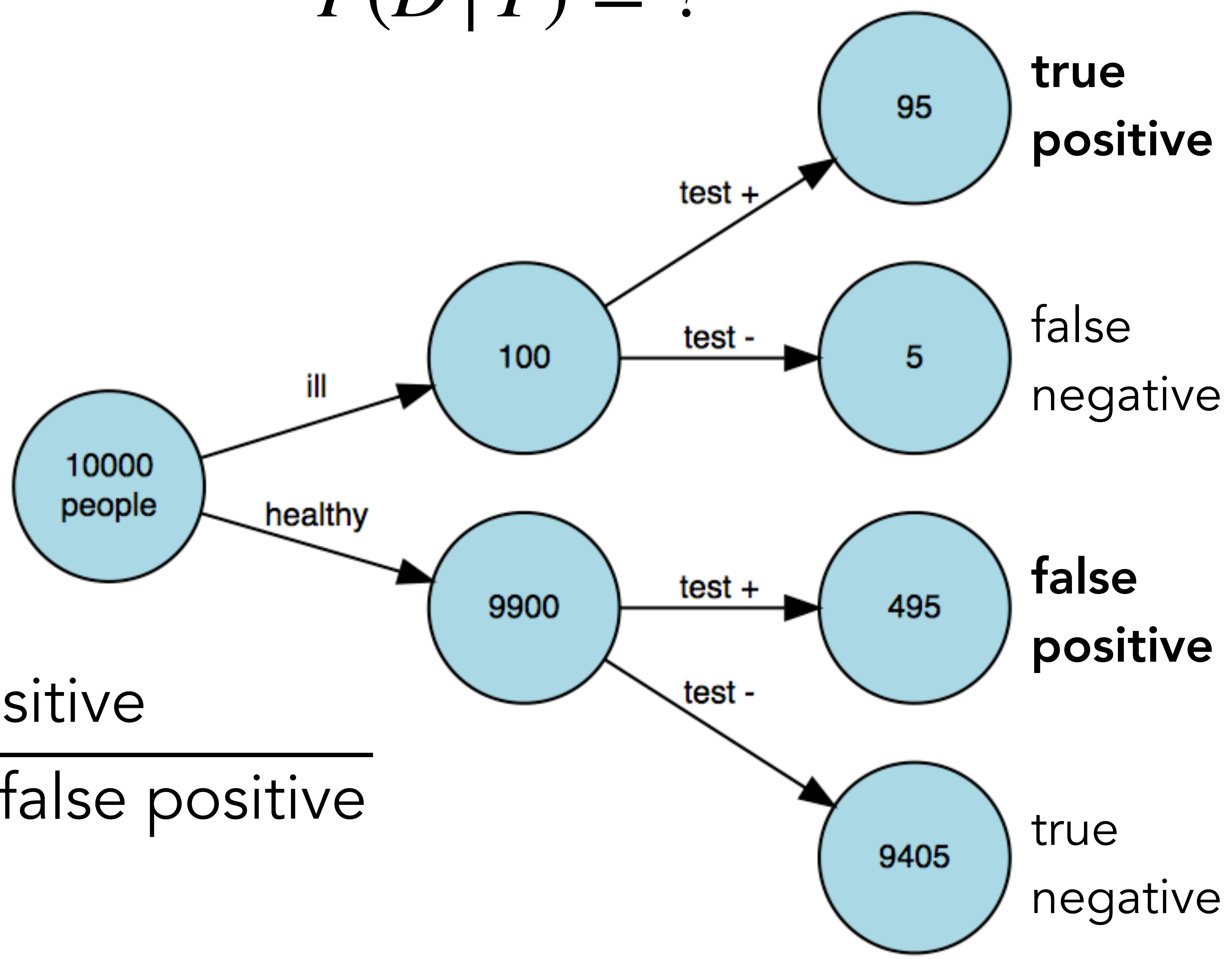
$$P(D) = 0.01$$

$$P(T \mid D) = 0.95$$

$$P(T \mid \neg D) = 0.05$$

what we want to know

$$P(D \mid T) = ?$$



$$P(D \mid T) = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$= \frac{95}{95 + 495}$$

$$\approx 0.16$$

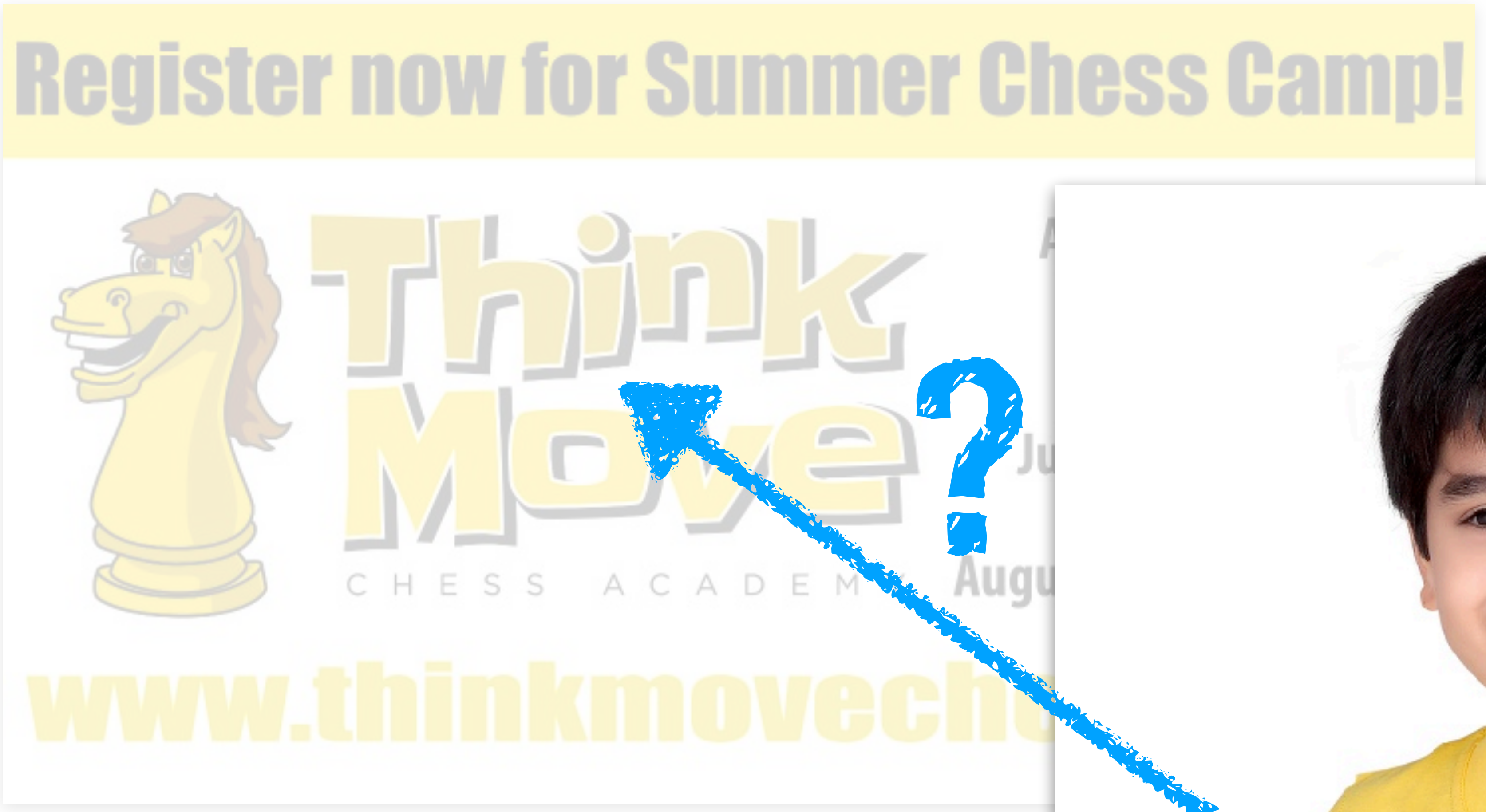
Summer camp

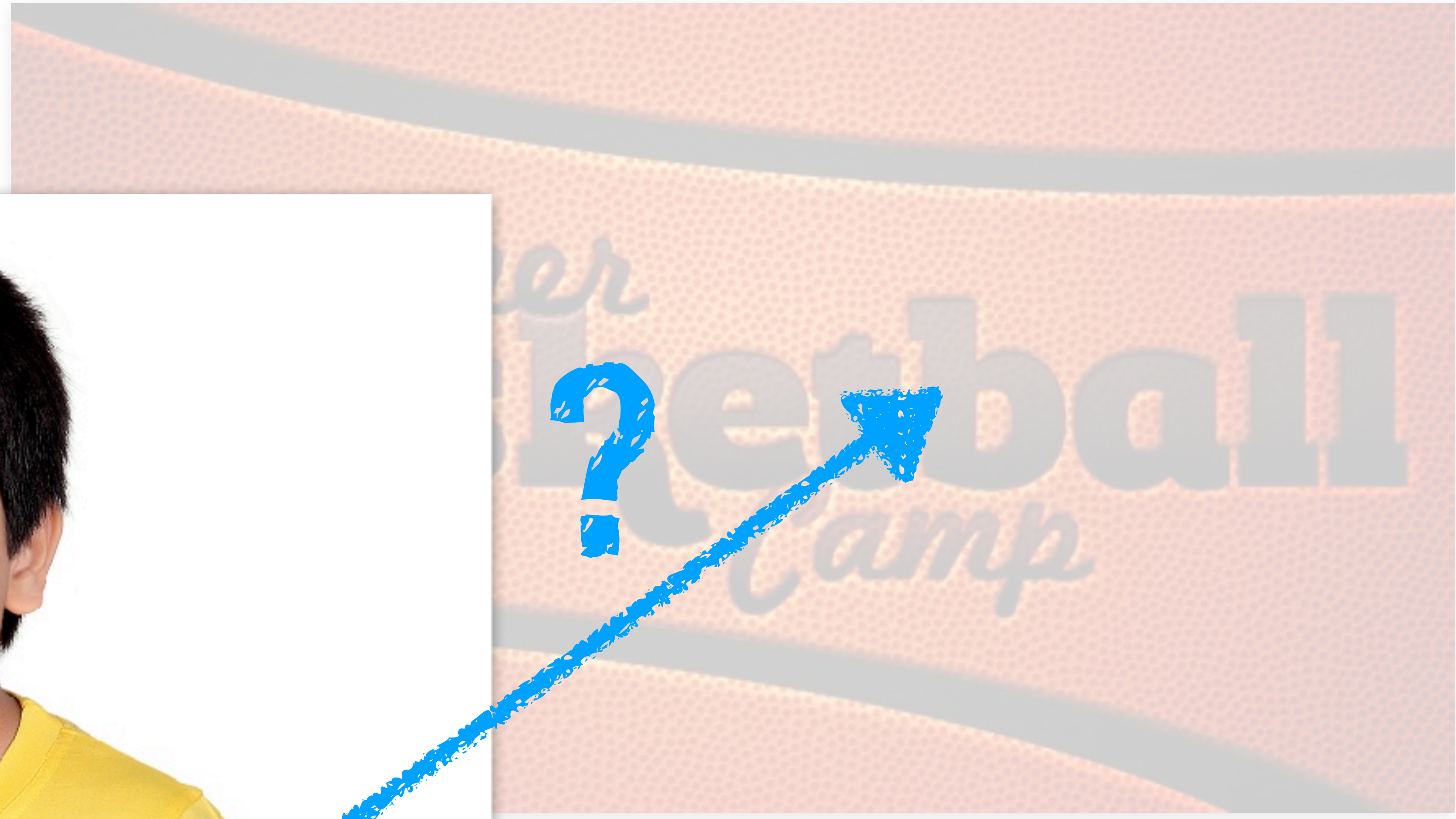
twice as many kids go to the basketball camp

$X \sim$ Normal$(\mu = 170, \sigma = 8)$

$X \sim$ Normal$(\mu = 180, \sigma = 10)$

Summer camp





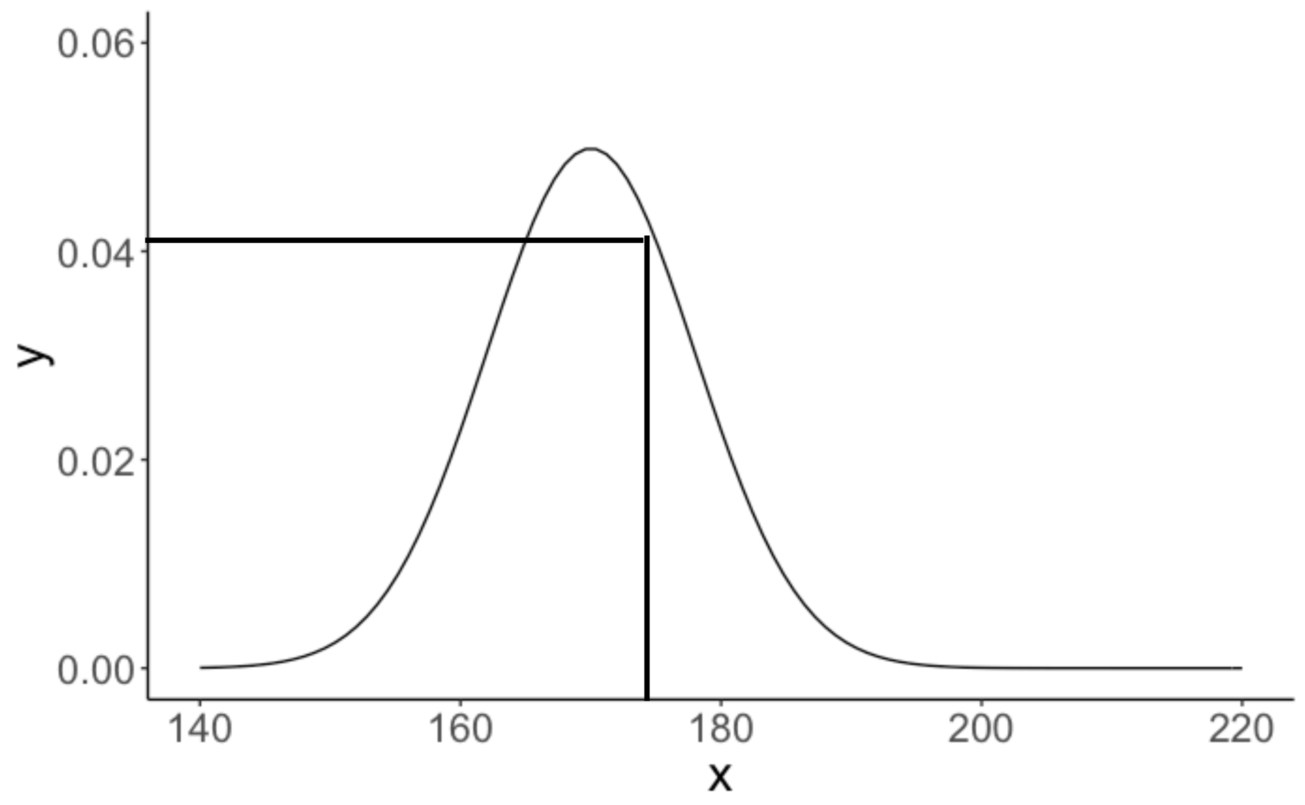
twice as mar … etball camp

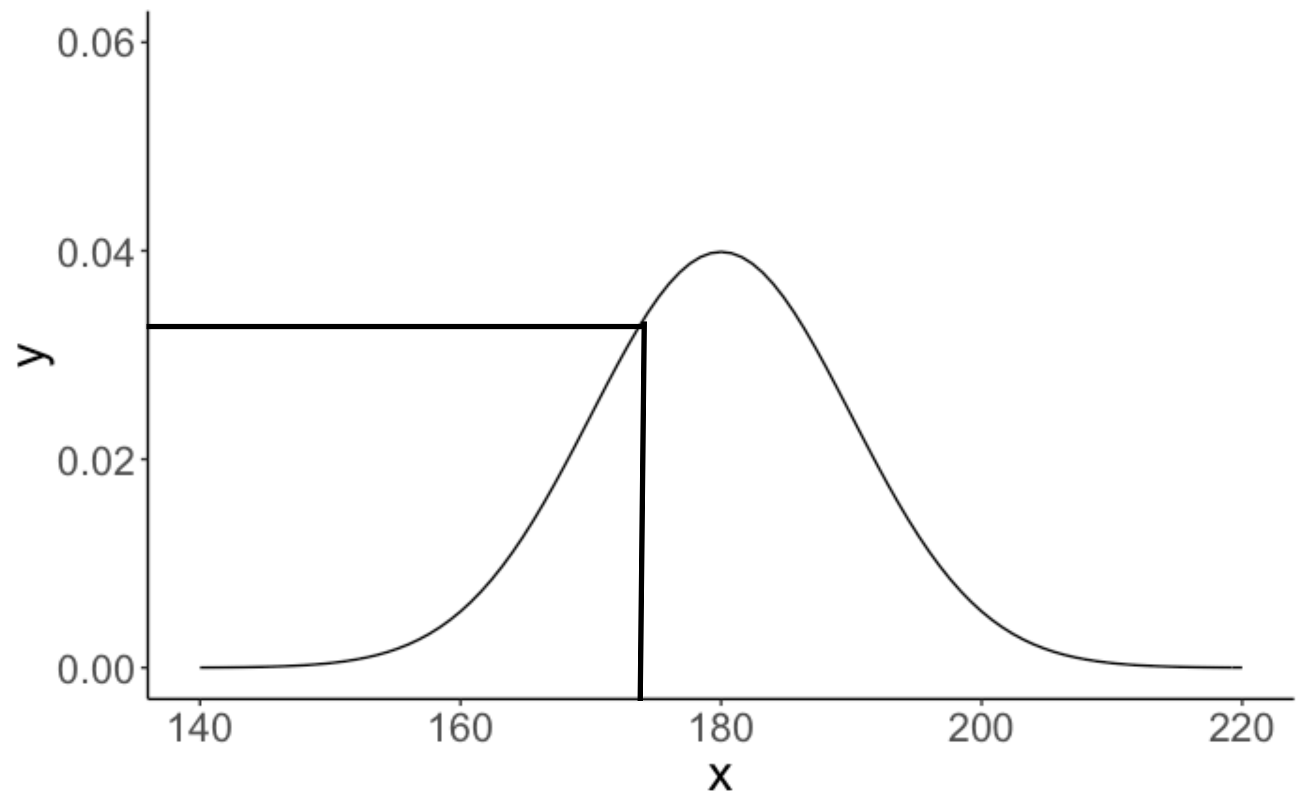
$X \sim$ Normal($\mu = 170$ … al($\mu = 180$, $\sigma = 10$)

Summer camp

prior $p(\text{chess}) = \frac{1}{3}$ $p(\text{basketball}) = \frac{2}{3}$

likelihood





`dnorm(175, mean = 170, sd = 8)` $= 0.041$

`dnorm(175, mean = 180, sd = 10)` $= 0.035$

posterior

$$p(\text{basketball} \mid 175) = \frac{p(175 \mid \text{basketball}) \cdot p(\text{basketball})}{p(175 \mid \text{basketball}) \cdot p(\text{basketball}) + p(175 \mid \text{chess}) \cdot p(\text{chess})}$$

$$p(\text{basketball} \mid 175) = \frac{0.035 \cdot 2/3}{0.035 \cdot 2/3 + 0.041 \cdot 1/3} \approx 0.63$$

send the kid to the basketball gym!

Flipping coins

Flipping coins

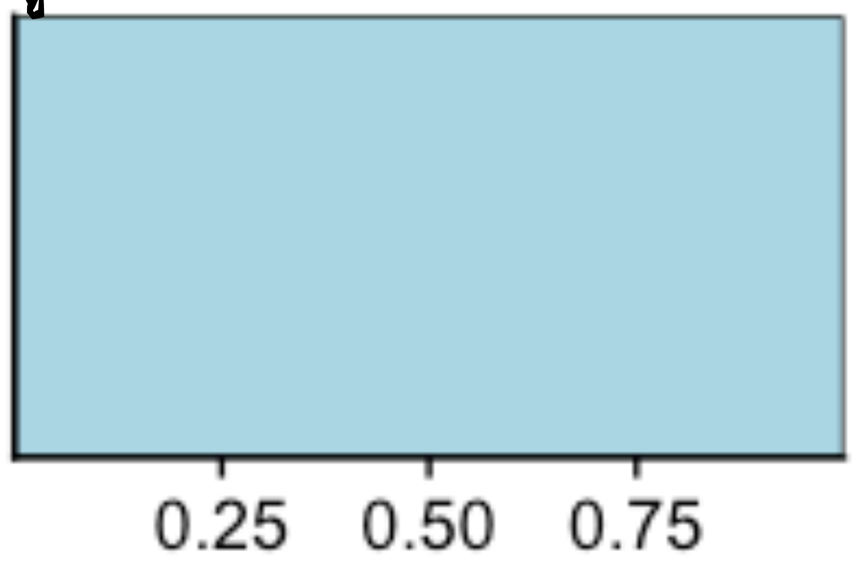
What's the probability of my coin coming up heads?

$$\theta = ?$$

Learning from data

How does/should our belief change as evidence comes in?

prior



Today's posterior is tomorrow's prior.

$$p(\theta \mid \mathrm{n}_{\mathrm{success}} = 6,\ \mathrm{n}_{\mathrm{trials}} = 8)$$

Coin flip example

Which coin did I flip?

Hypotheses

$\theta = 0.3$

$\theta = 0.5$

$\theta = 0.9$

Data

#8 tails, #2 heads

Bayesian Recipe

- Hypotheses
- Prior over hypotheses
- Data
- Likelihood of the data given each hypothesis
- Posterior over hypotheses given the data

+ a healthy dose of Bayes' rule

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

Coin flip example

```r
# data
data = rep(0:1, c(8, 2))

# parameters
theta = c(0.1, 0.5, 0.9)

# prior
prior = c(0.25, 0.5, 0.25)

# likelihood
likelihood = dbinom(sum(data == 1), size = length(data), prob = theta)

# posterior
posterior = likelihood * prior / sum(likelihood * prior)
```

8 tails and 2 heads

hypotheses

prior over the hypotheses

binomial distribution

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

law of total probability:

$$p(D) = \sum_{i=1}^{n} p(D|H_i) \cdot p(H_i)$$

Coin flip example

```r
# data
data = rep(0:1, c(8, 2))

# parameters
theta = c(0.1, 0.5, 0.9)

# prior
prior = c(0.25, 0.5, 0.25)

# likelihood
likelihood = dbinom(sum(data == 1), size = length(data), prob = theta)

# posterior
posterior = likelihood * prior / sum(likelihood * prior)
```

8 tails and 2 heads

hypotheses

prior over the hypotheses

binomial distribution

multiply

re-normalize

theta	prior	likelihood	prior_x_like	posterior
0.1	0.25	0.19	0.0475	0.69
0.5	0.50	0.04	0.02	0.31
0.9	0.25	0.00	0.00	0.00

Coin flip example

data: #8 tails, #2 heads

Which coin was flipped?

what the model knows **before** having seen the data

learning by **conditioning** on the data

what the model knows **after** having seen the data

p = 0.1

p = 0.5

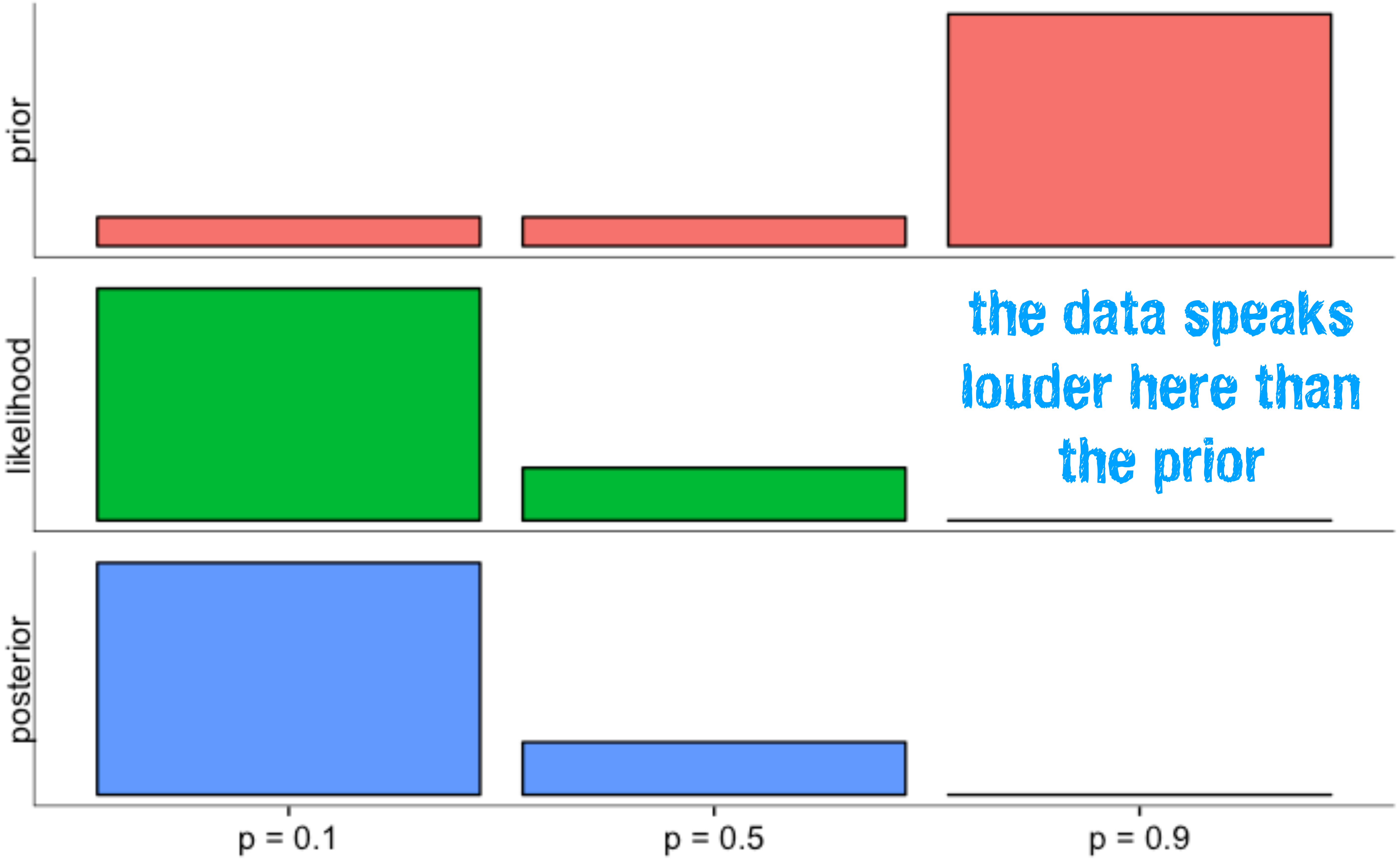
p = 0.9

posterior = multiplicative weighting of prior and likelihood

Coin flip example

data: #8 tails, #2 heads

Which coin was flipped?

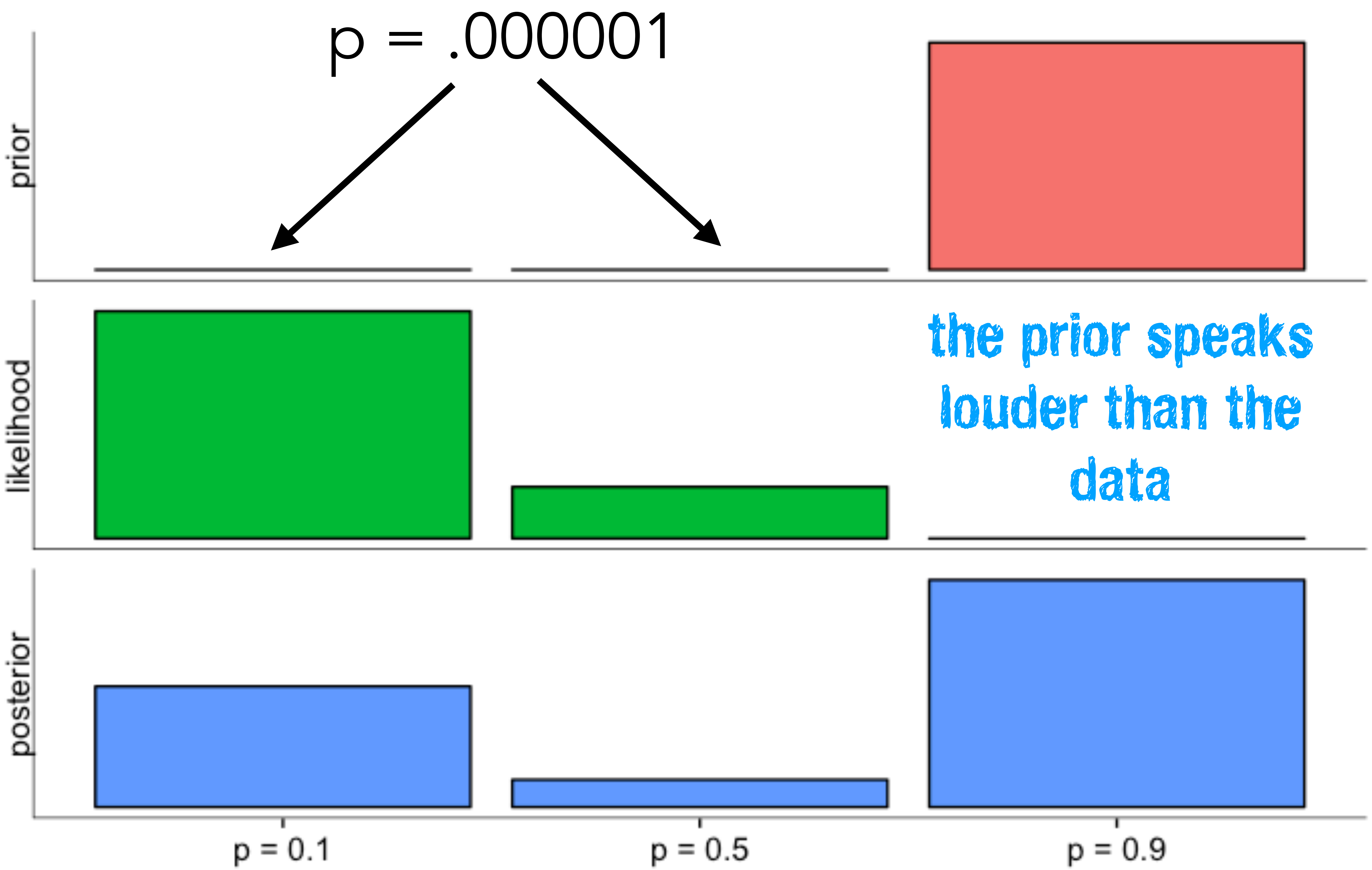


posterior = multiplicative weighting of prior and likelihood

Coin flip example

data: #8 tails, #2 heads

Which coin was flipped?



posterior = multiplicative weighting of prior and likelihood

What affects the posterior?

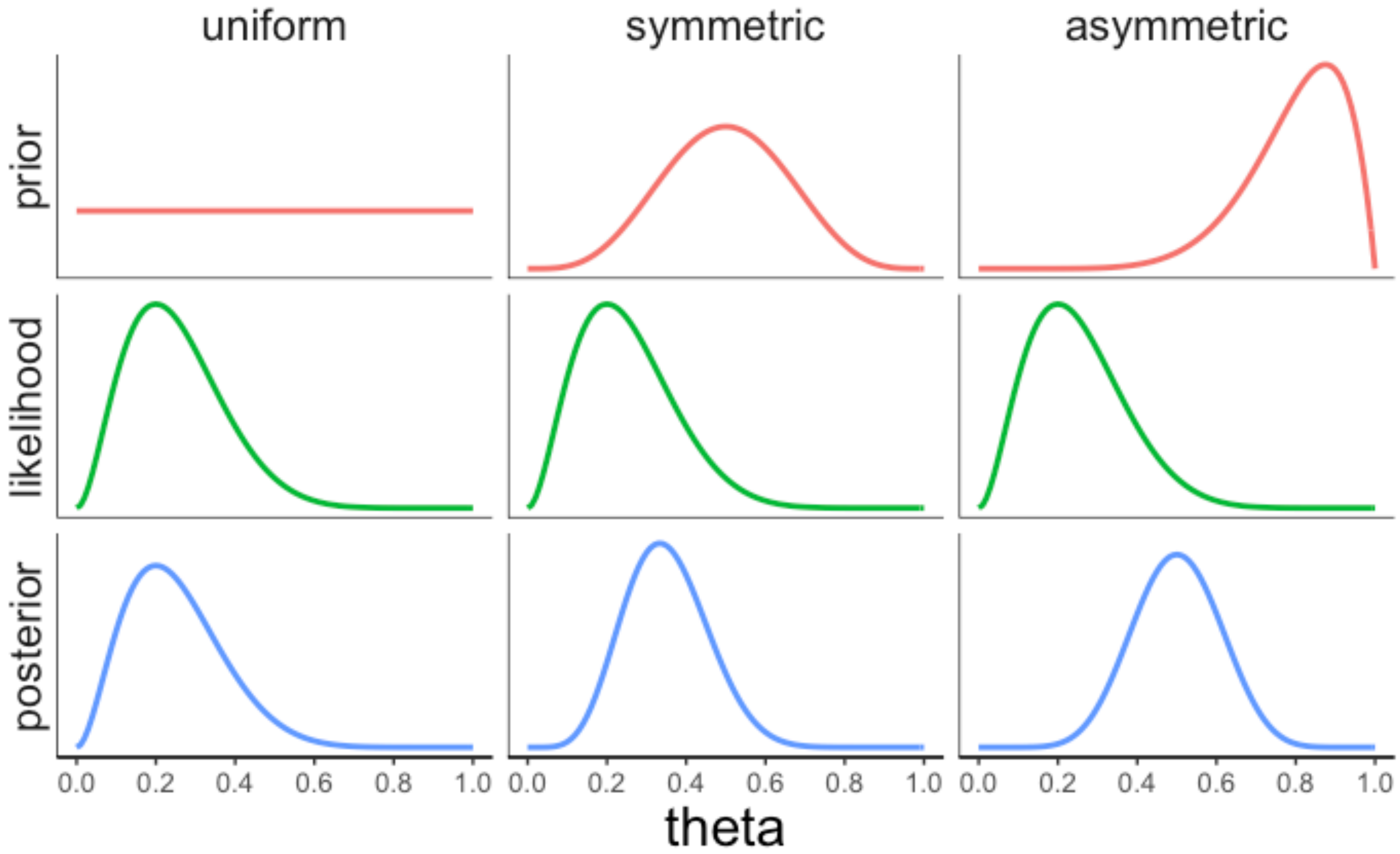
What affects the posterior?

1. the prior over hypotheses
2. the likelihood of the data given each hypothesis

$$\underbrace{p(H|D)}_{\text{Posterior}} = \frac{\overbrace{p(D|H)}^{\text{Likelihood}} \cdot \overbrace{p(H)}^{\text{Prior}}}{\underbrace{p(D)}_{\text{Normalizing constant}}}$$

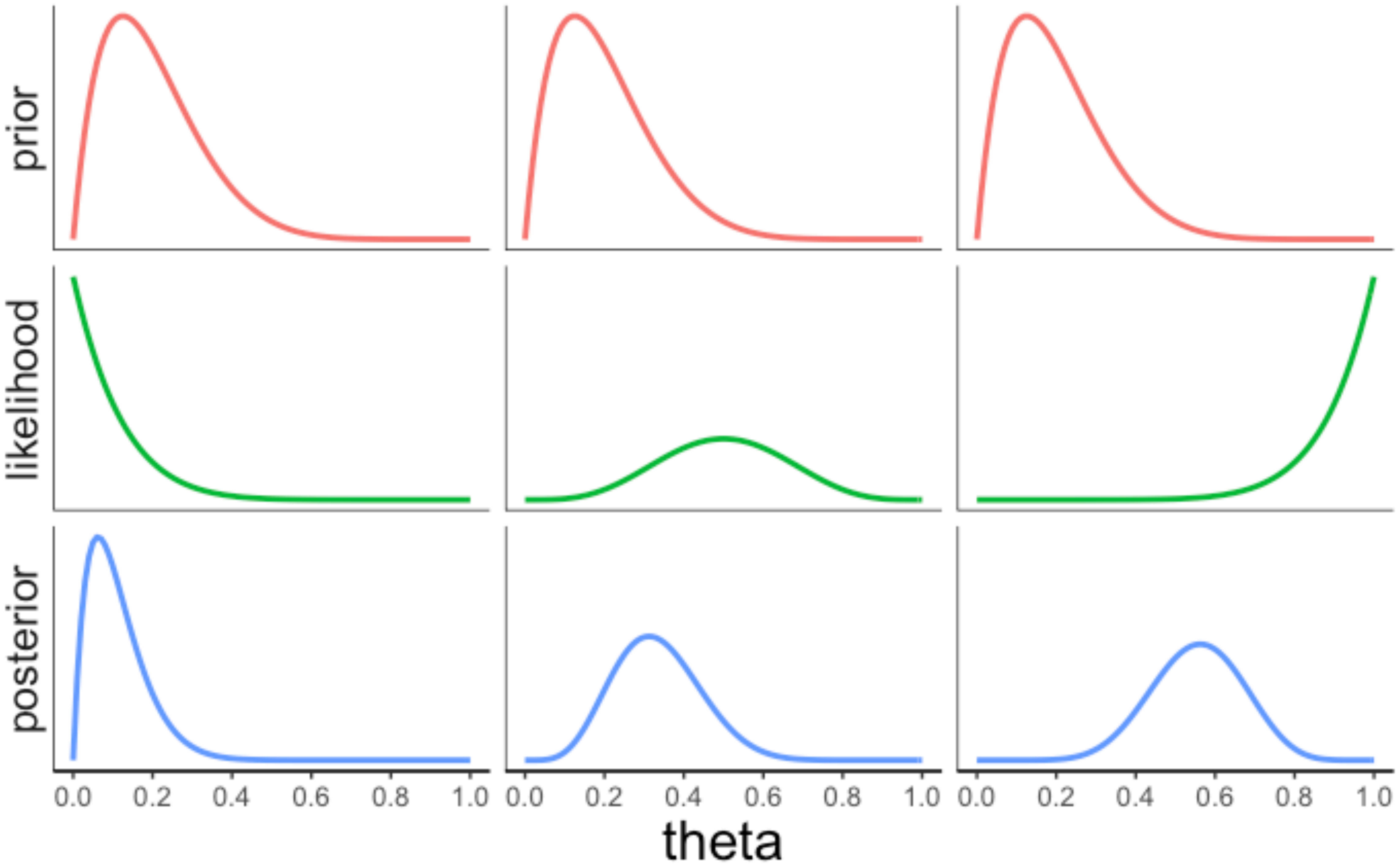
The effect of the **prior**

same data, different priors

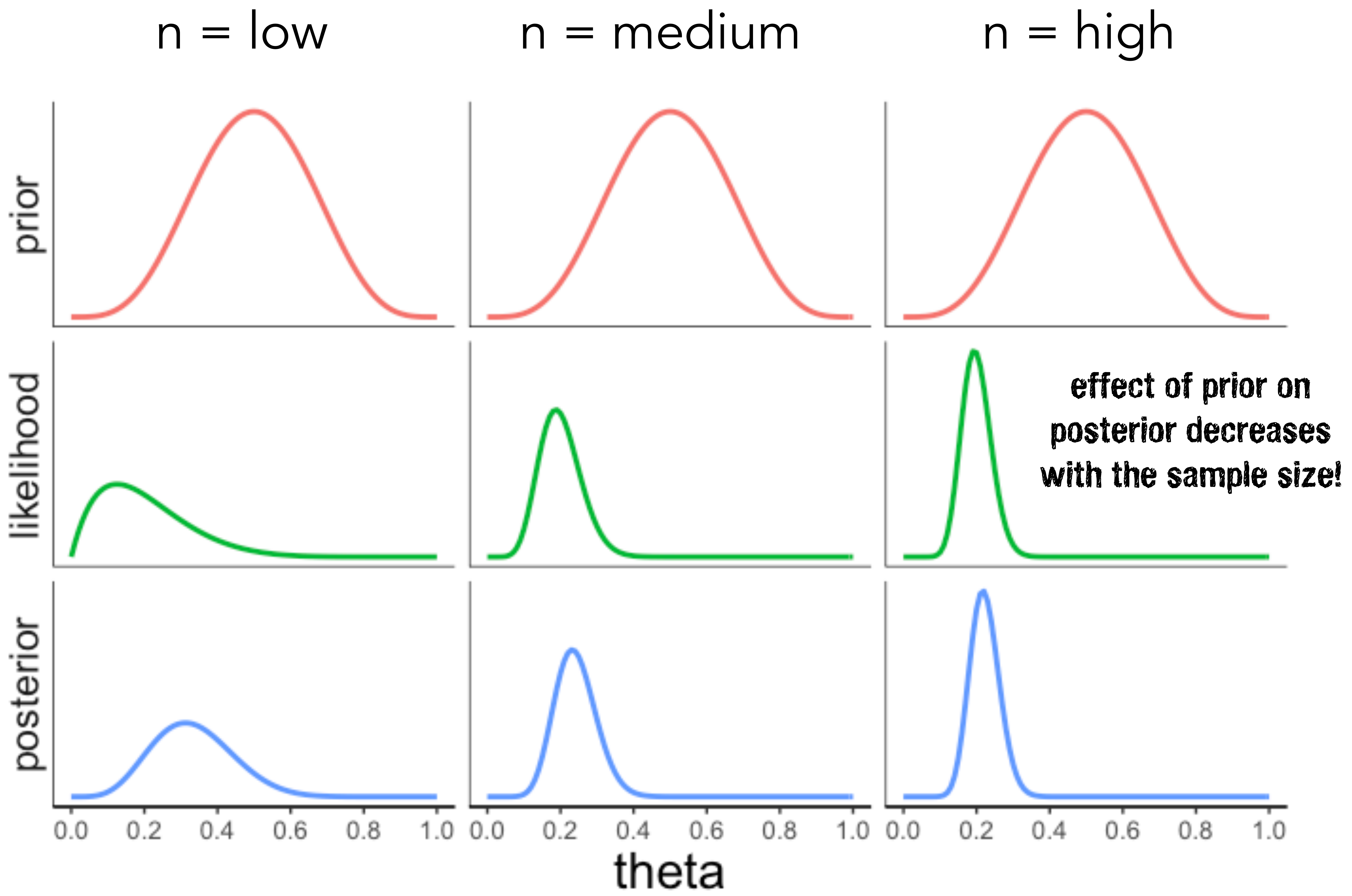


The effect of the **likelihood**

same prior, different data



The effect of **sample size**



Ingredients: likelihood, prior, inference

Ingredients

$$\underbrace{p(H|D)}_{\text{Posterior}} = \frac{\overbrace{p(D|H)}^{\text{Likelihood}} \cdot \overbrace{p(H)}^{\text{Prior}}}{\underbrace{p(D)}_{\text{Normalizing constant}}}$$

Likelihood Prior

Posterior

$$p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{p(D)}$$

Normalizing constant

Likelihood

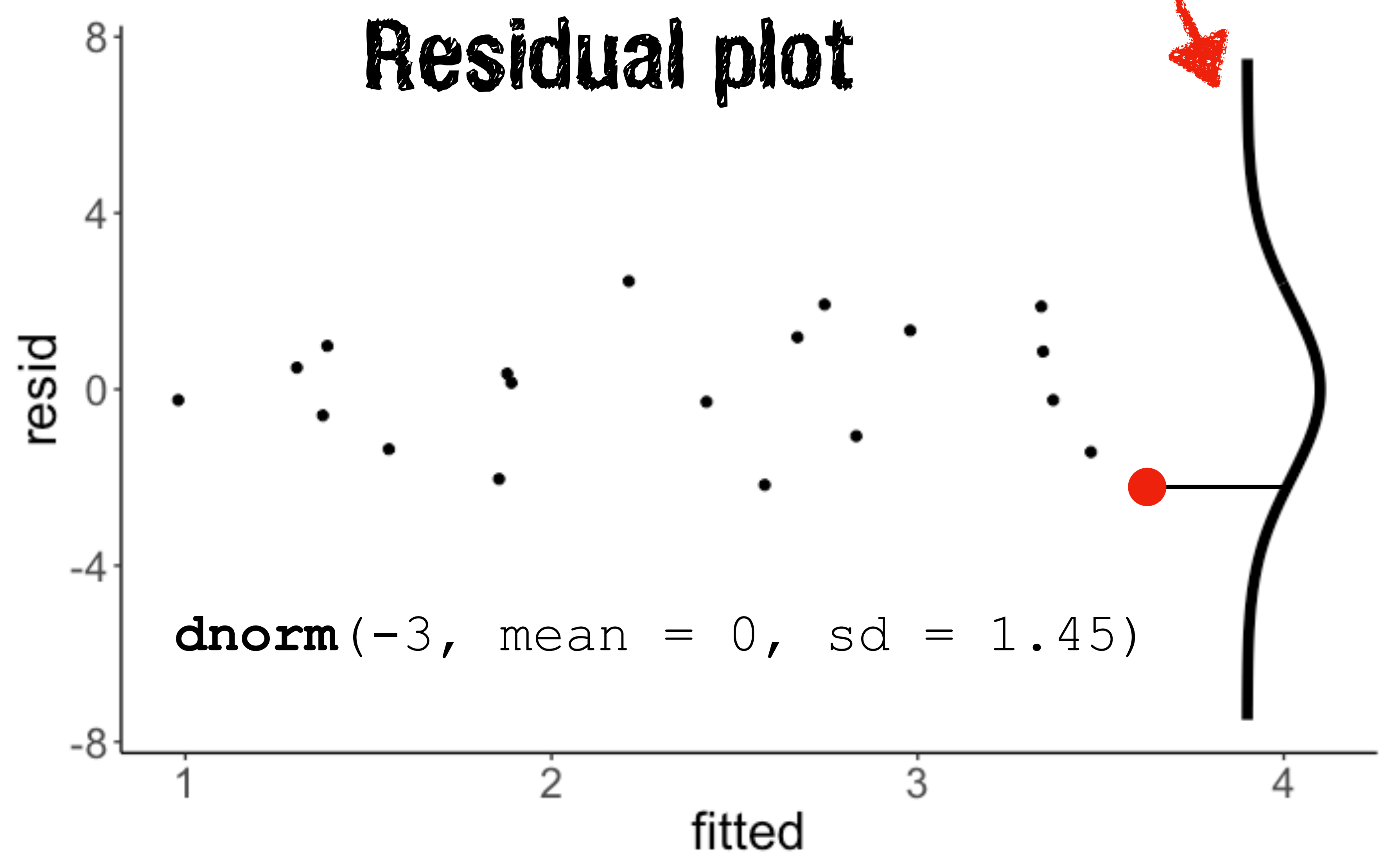
- **What probabilistic model describes best how the data were generated?**
  - What assumptions can you make about the data?
    - What's the nature of your dependent variable (e.g. binary, ordered, continuous)?
  - Does the model re-create the behavior of interest?

Likelihood

Gaussian distribution

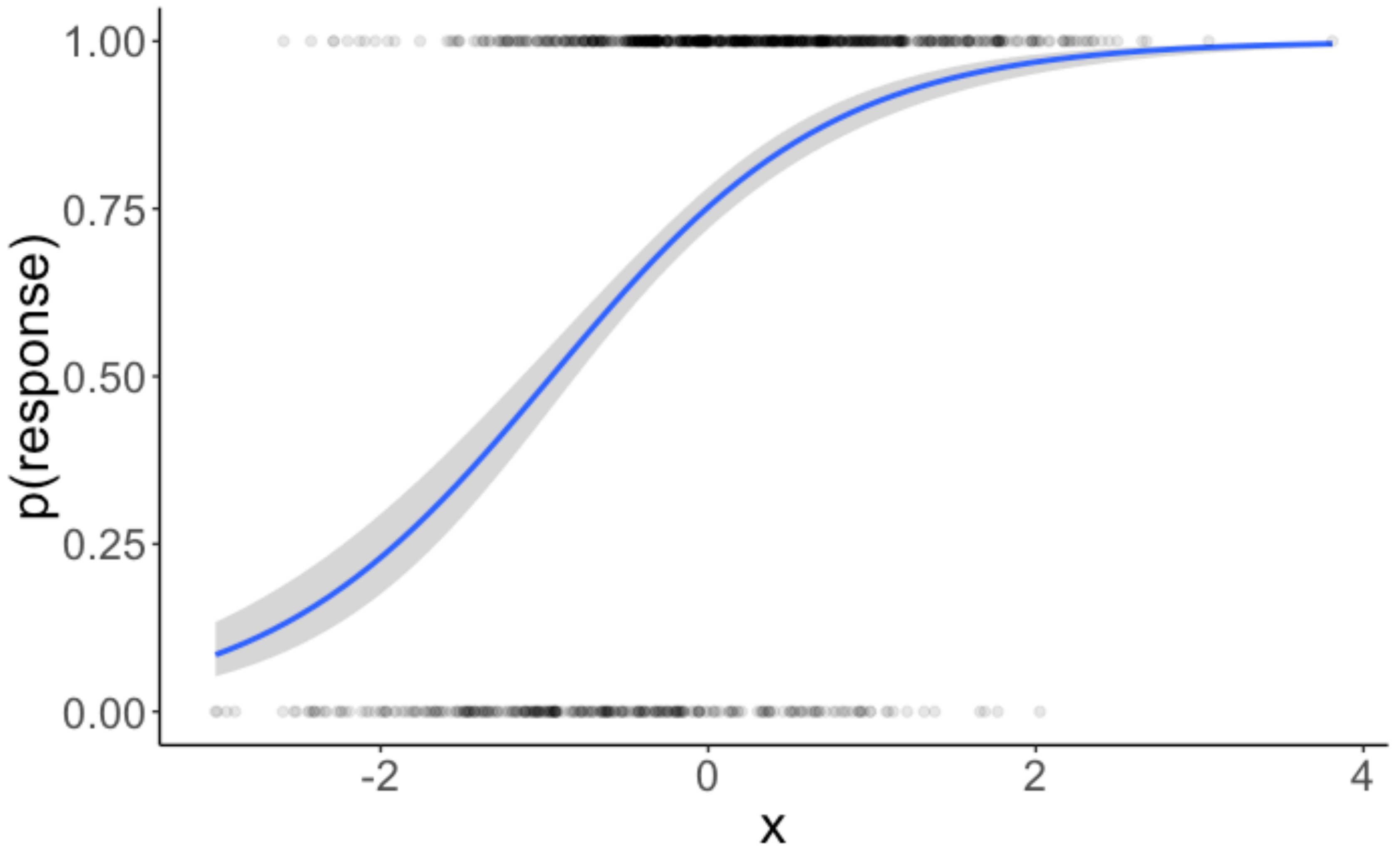
$$Y_i = b_0 + b_1 \cdot x_i + e_i$$

$$e_i \sim \mathcal{N}(0, \sigma)$$

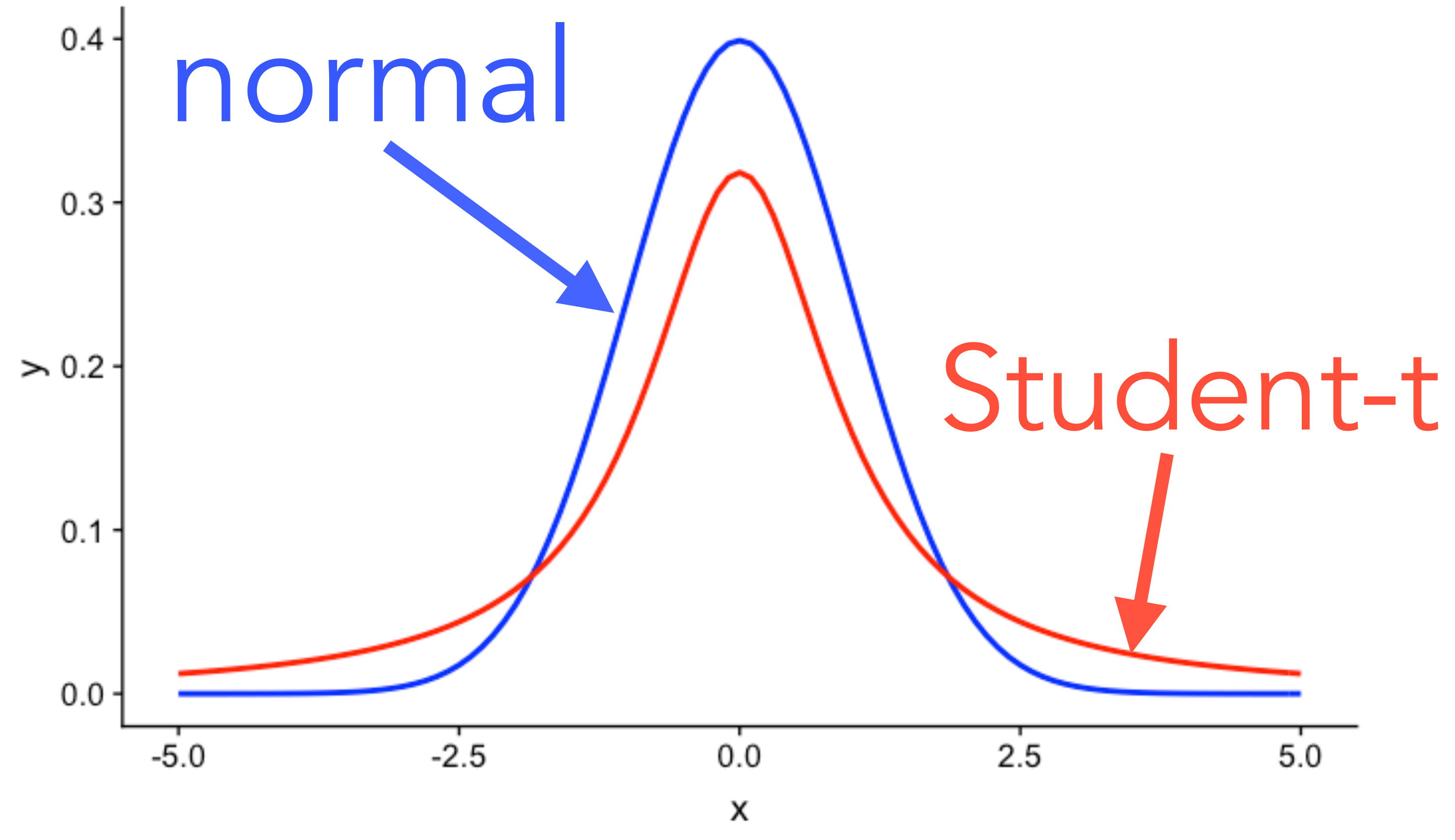


Binomial distribution

```
fit.glm = glm(formula = survived ~ 1 + fare,
              family = "binomial",
              data = df.titanic)
```

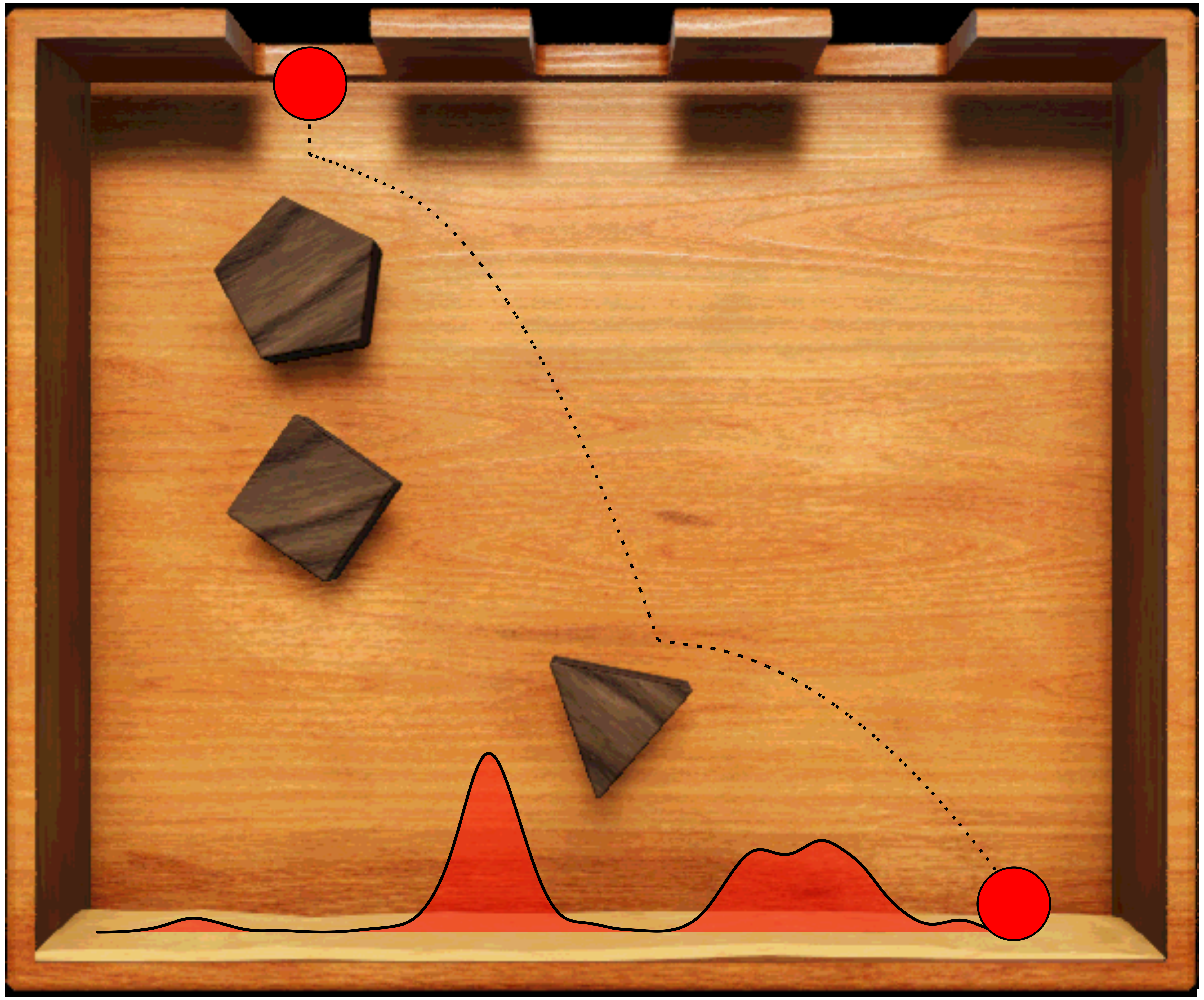


Likelihood



- **Bernoulli**:
  - binary data
  - a single trial
- **Poisson**: count of discrete events
- **Beta-binomial**: like binomial but probability of success may change across trials
- **Student-*t***:
  - same as Normal
  - handles greater variability in the data (distribution has **fat tails**)
- …

Prediction: Where will the ball land?

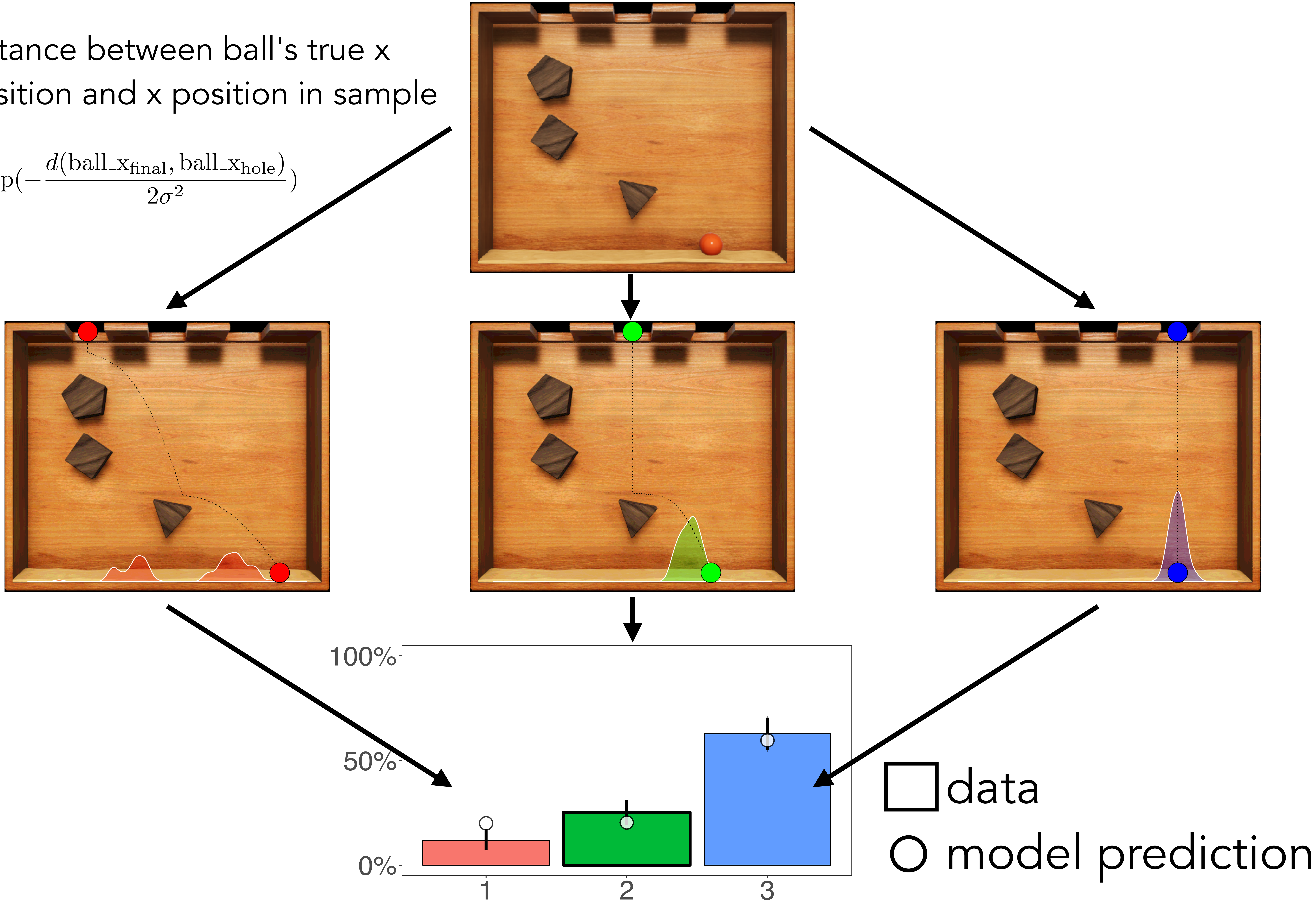


Aggregated responses

Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

$$\exp\left(-\frac{d(\text{ball\_x}_{\text{final}}, \text{ball\_x}_{\text{hole}})}{2\sigma^2}\right)$$



Prior

Likelihood Prior

Posterior

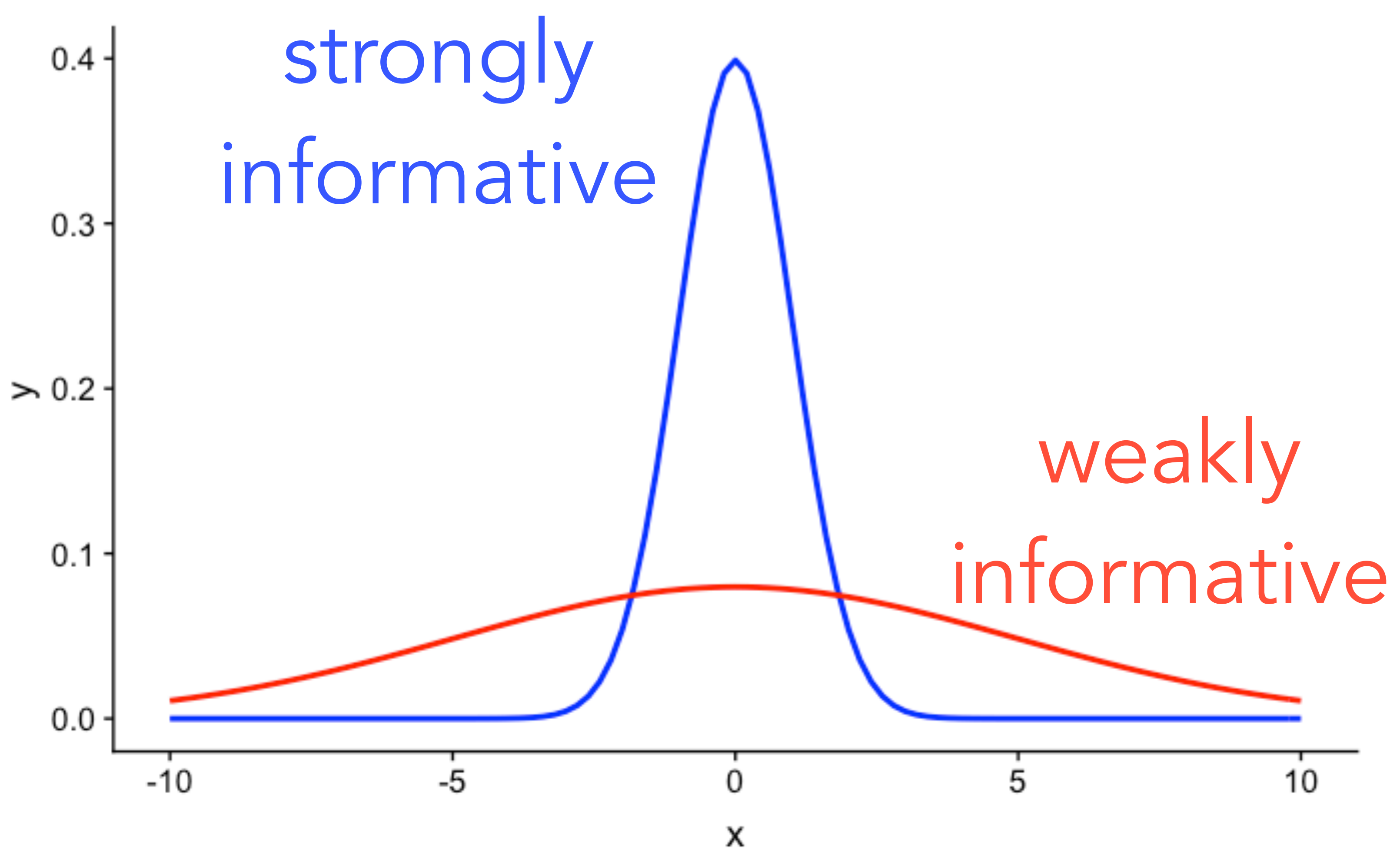
$$p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{p(D)}$$

Normalizing constant

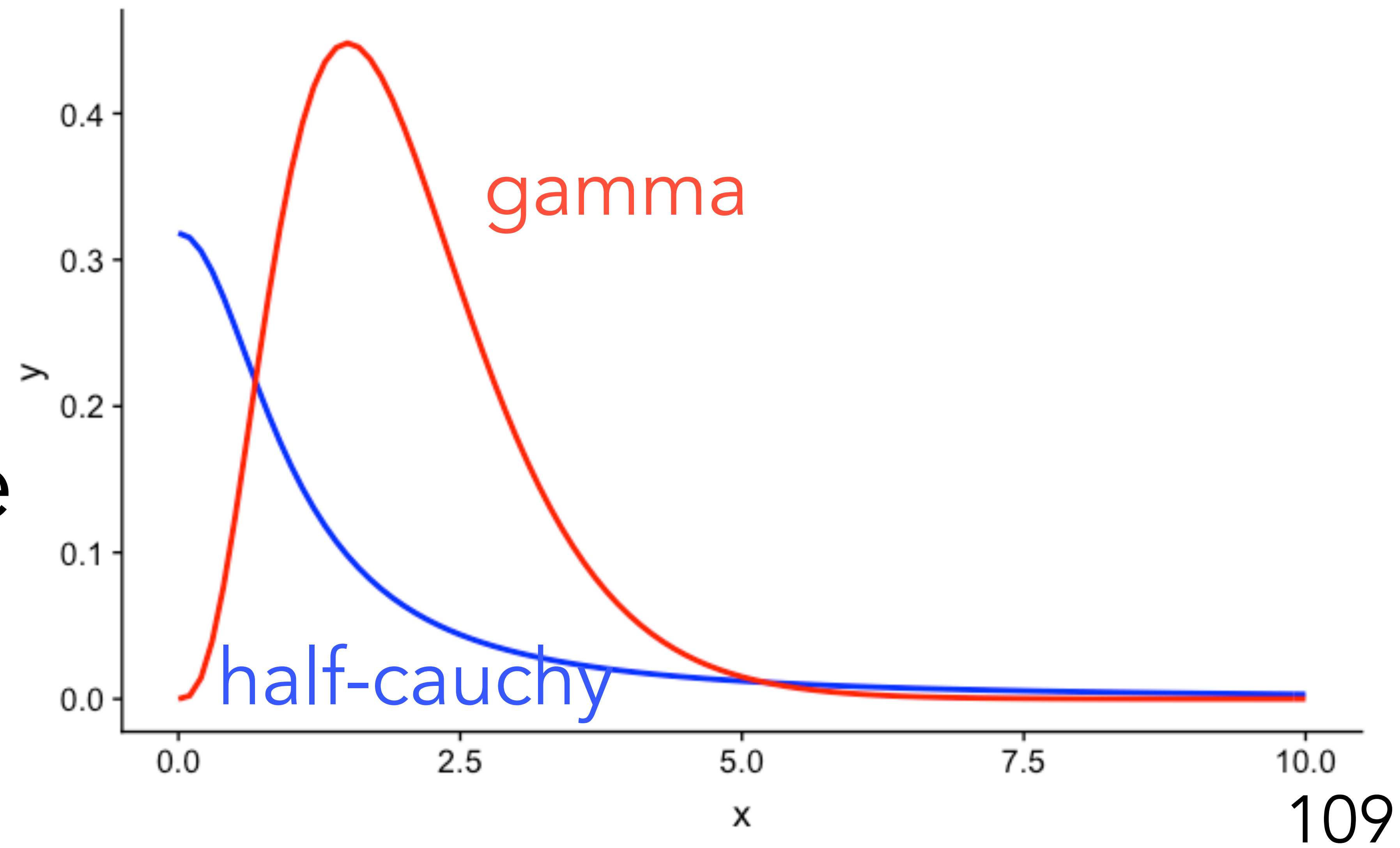
Prior

- **uniform**:
  - continuous or discrete
  - bounded between minimum and maximum
- **gaussian:**
  - sd determines how informative the prior is
- **gamma, half-cauchy**:
  - for parameters we know are positive

for **beta coefficients** in a regression



for **standard deviation** of the Gaussian



Inference

Likelihood Prior

Posterior

$$\underbrace{p(H|D)}_{\text{Posterior}} = \frac{\overbrace{p(D|H)}^{\text{Likelihood}} \cdot \overbrace{p(H)}^{\text{Prior}}}{\underbrace{p(D)}_{\text{Normalizing constant}}}$$

Normalizing constant

the devil is in the denominator ...

Doing Bayesian inference

Discrete hypothesis space

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{\sum_{i=1}^{n} p(D|H_i) \cdot p(H_i)}$$

sum over all possibilities

Continuous hypothesis space

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{\int_{-\infty}^{\infty} p(D|H_i) \cdot p(H_i) dH_i}$$

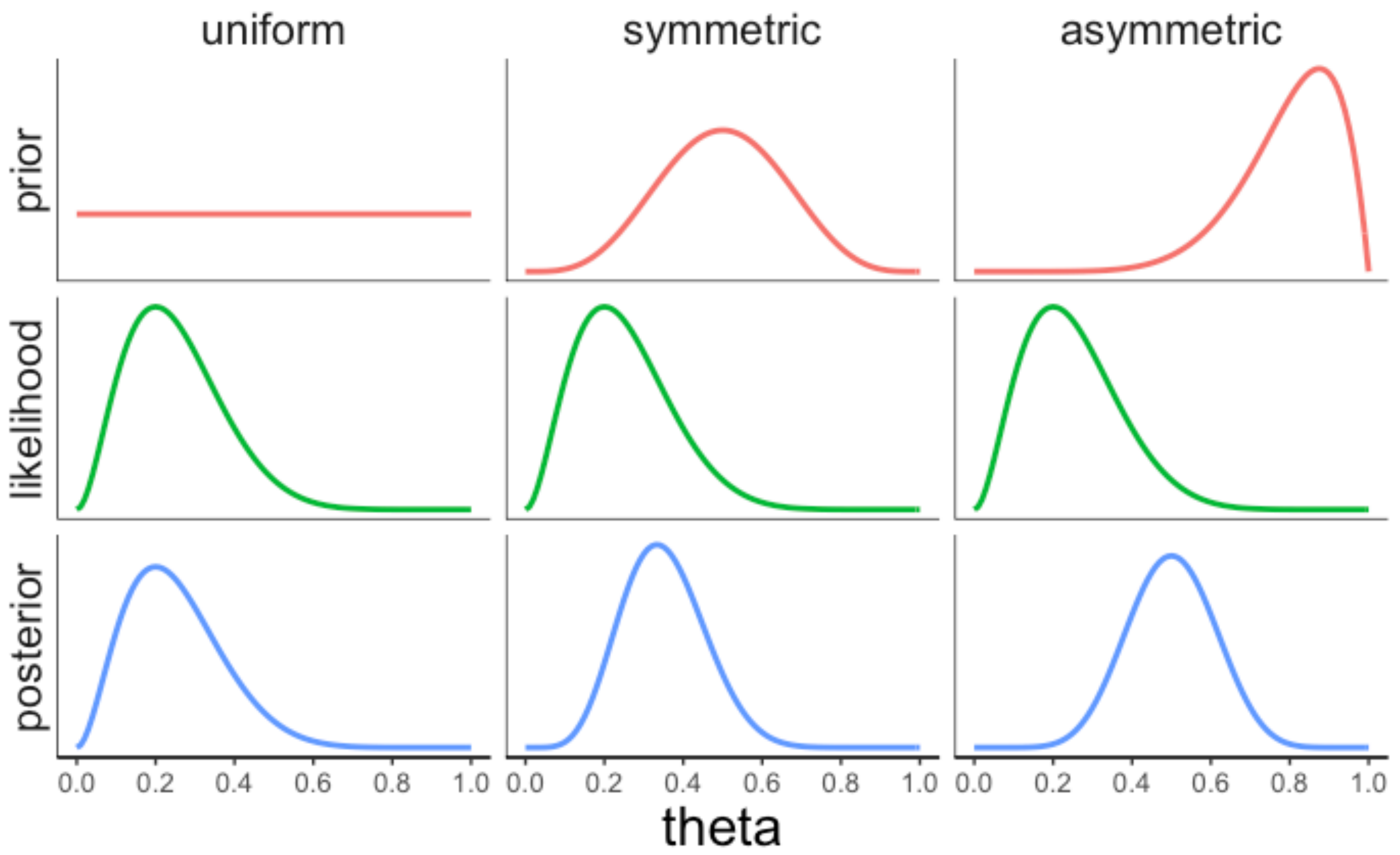
integral over all possibilities

Discretizing the parameters

```r
# grid
theta = seq(0, 1, 0.01)

# data
data = rep(0:1, c(8, 2))

# calculate posterior
df.prior = tibble(theta = theta,
                  prior_uniform = dbeta(theta, shape1 = 1, shape2 = 1),
                  prior_normal = dbeta(theta, shape1 = 5, shape2 = 5),
                  prior_biased = dbeta(theta, shape1 = 8, shape2 = 2)) %>%
  pivot_longer(cols = -theta,
               names_to = "prior_index",
               values_to = "prior") %>%
  mutate(likelihood = dbinom(sum(data == 1),
                             size = length(data),
                             prob = theta)) %>%
  group_by(prior_index) %>%
  mutate(posterior = likelihood * prior / sum(likelihood * prior)) %>%
  ungroup() %>%
  pivot_longer(cols = -c(theta, prior_index),
               names_to = "index",
               values_to = "value")
```

100 discrete values

for 3 variables, we would already need 1 Mio combinations



The CURSE of dimensionality

Inference via sampling

- we cannot directly calculate the probability of the posterior (because it might have a pretty weird shape)
- **but**: we can draw random samples from the posterior
- we can then use our data wrangling and visualization skills to make inferences based on these samples

It's as if ...

we don't have **`pnorm`**`()` but we do have **`rnorm`**`()`

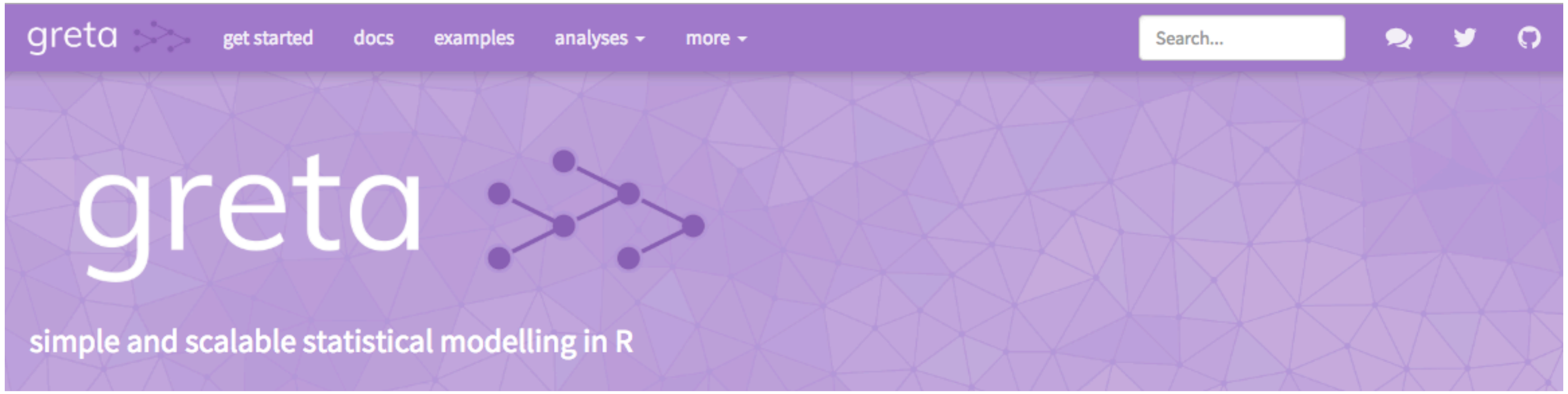
Inference via sampling

- Bayesian data analysis is becoming more popular because:
  - computers are getting more powerful
  - inference techniques are getting better
  - software packages become easier to use

Doing Bayesian data analysis

Software packages

`library("greta")`

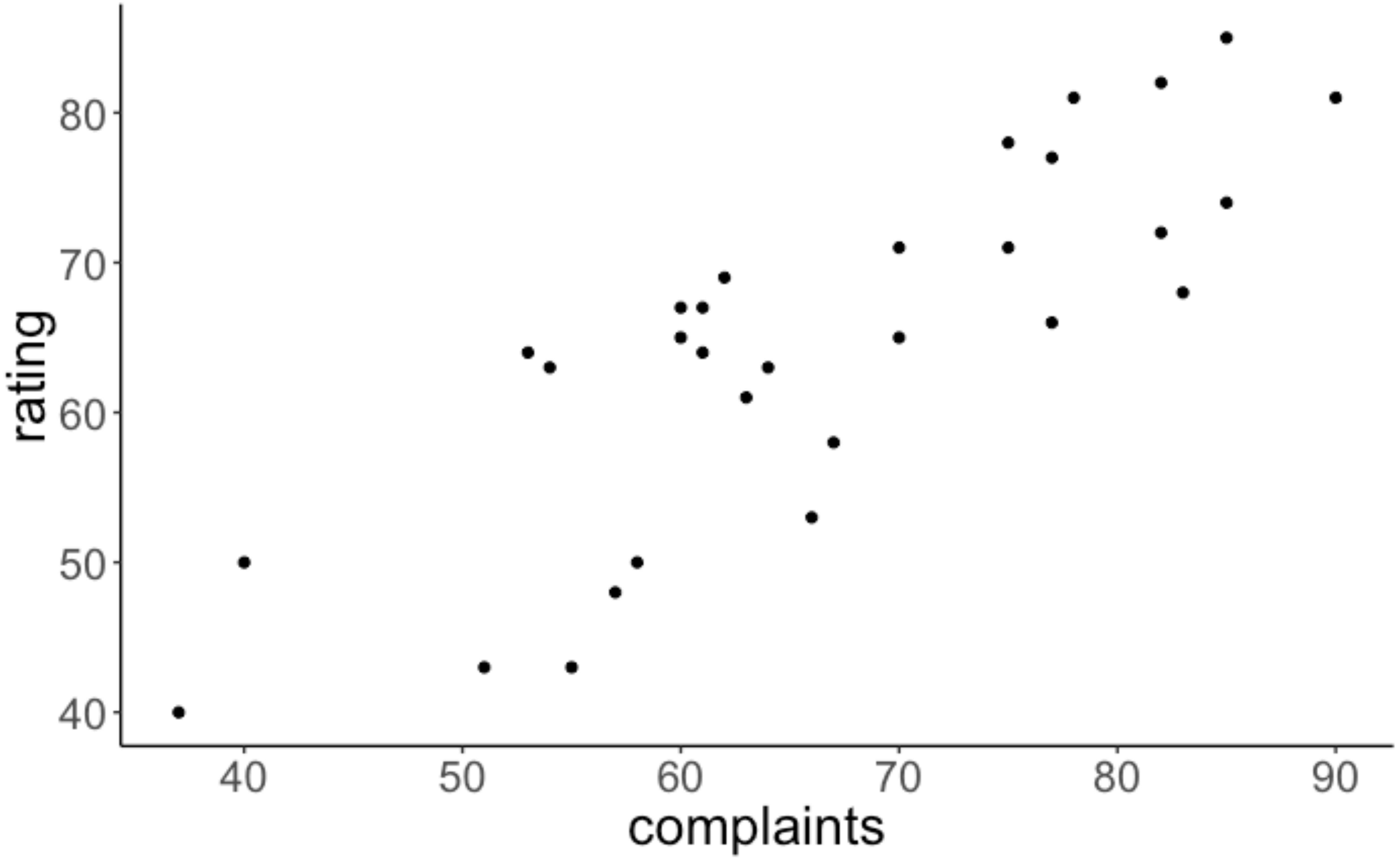


- let's us write Bayesian models directly in R with a simple syntax
- uses Tensorflow to implement Hamiltonian Monte Carlo sampling (a fast inference algorithm ...)

unfortunately doesn't work on Apple silicone (e.g. M1)

Attitude data set

What's the relationship between how well an employee handles complaints and their overall rating?



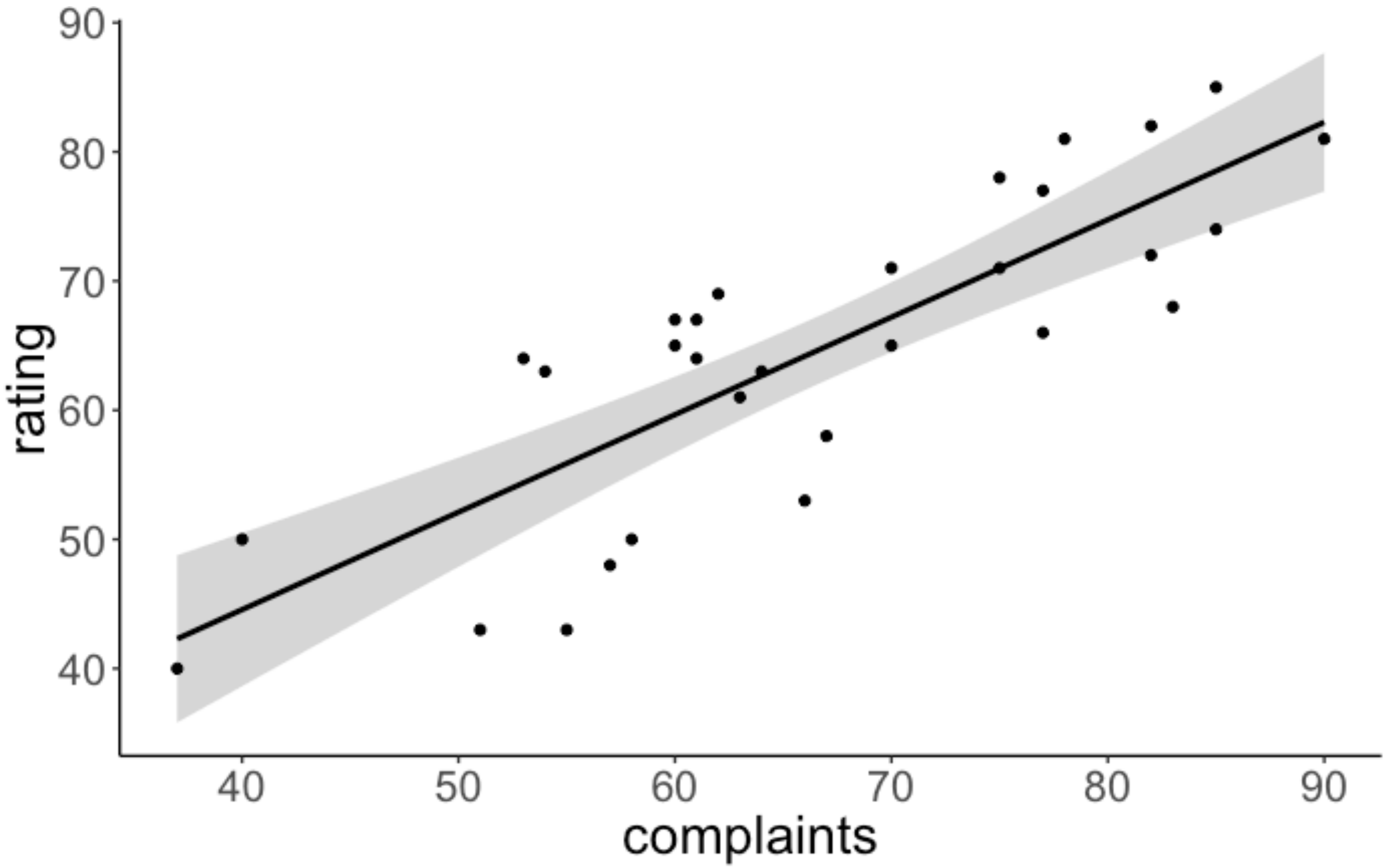
Frequentist analysis

Frequentist analysis

```r
# fit model
fit = lm(formula = rating ~ 1 + complaints,
         data = df.attitude)

# print summary
fit %>% summary()
```

```
Call:
lm(formula = rating ~ 1 + complaints, data = df.attitude)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.8799  -5.9905   0.1783   6.2978   9.6294 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 14.37632    6.61999   2.172   0.0385 *  
complaints   0.75461    0.09753   7.737 1.99e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.993 on 28 degrees of freedom
Multiple R-squared:  0.6813,    Adjusted R-squared:  0.6699 
F-statistic: 59.86 on 1 and 28 DF,  p-value: 1.988e-08
```

Visualize model predictions



Best-fitting regression line with confidence interval

Bayesian analysis

Model specification

```r
library("greta")
library("tidybayes")

# variables & priors
b0 = normal(0, 10)
b1 = normal(0, 10)
sd = cauchy(0, 3, truncation = c(0, Inf))

# linear predictor
mu = b0 + b1 * attitude$complaints

# observation model (likelihood)
distribution(attitude$rating) = normal(mu, sd)

# define the model
m = model(b0, b1, sd)
```

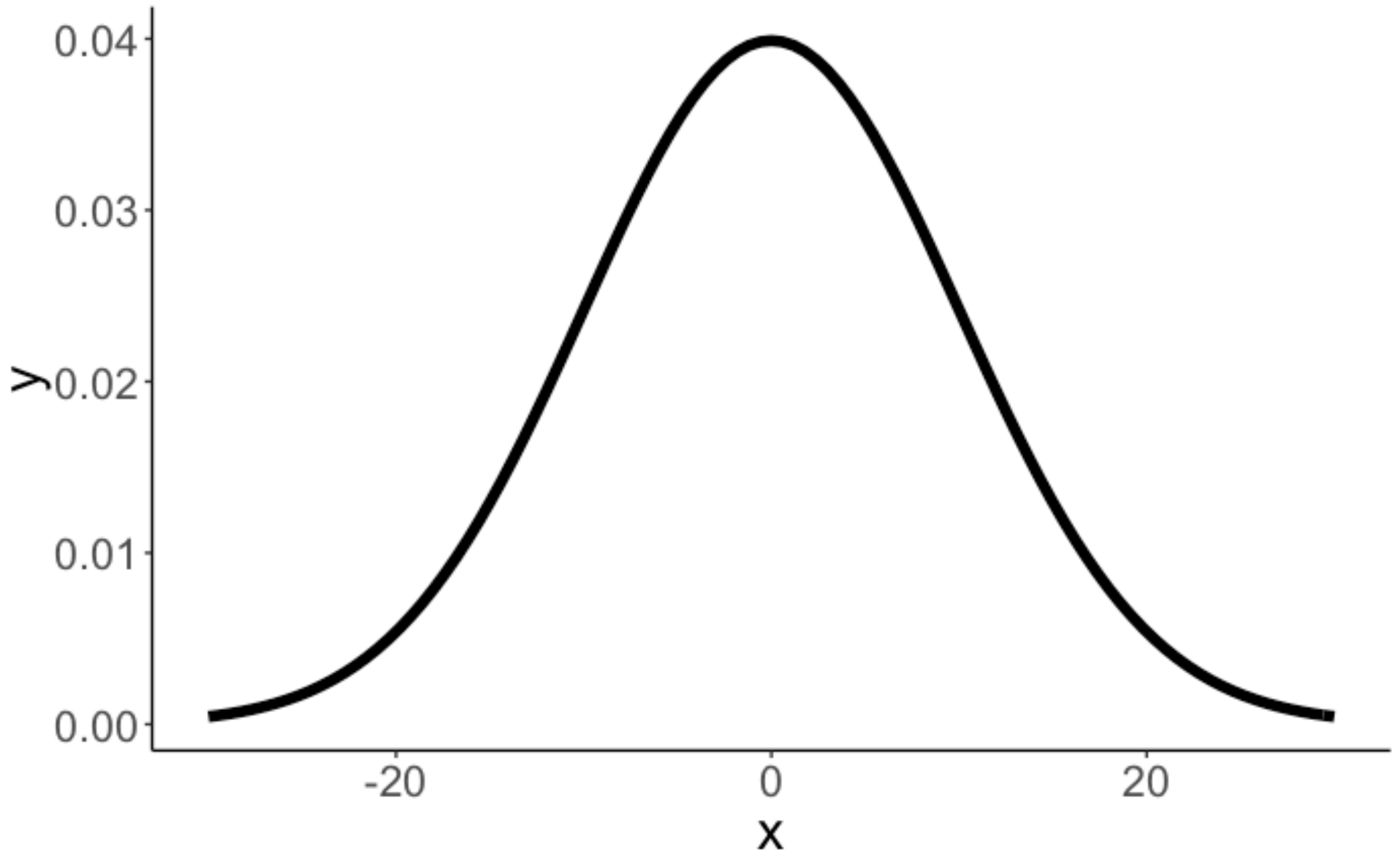
priors

linear combination

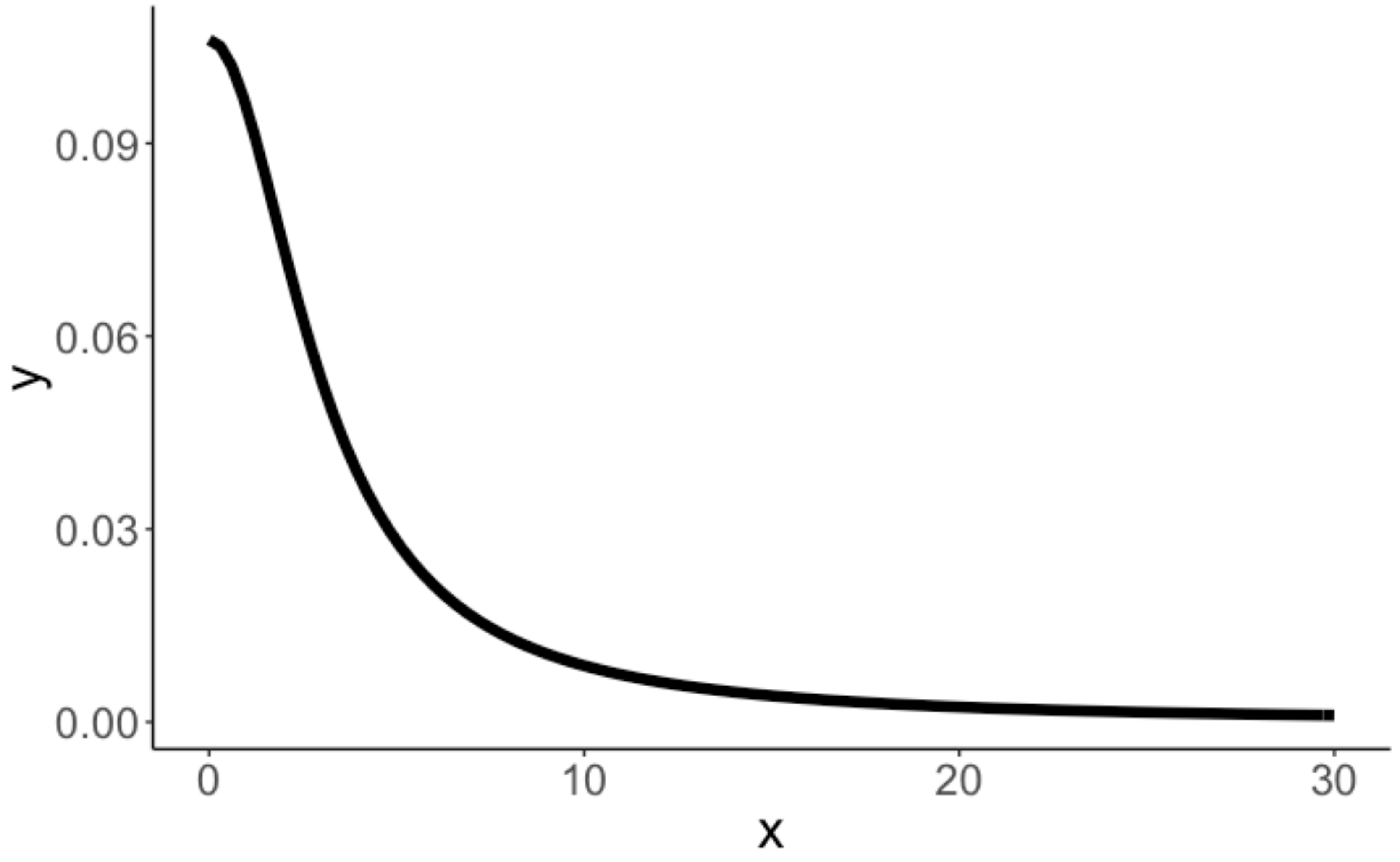
Gaussian likelihood

build the model

Priors



Gaussian prior on intercept and coefficient

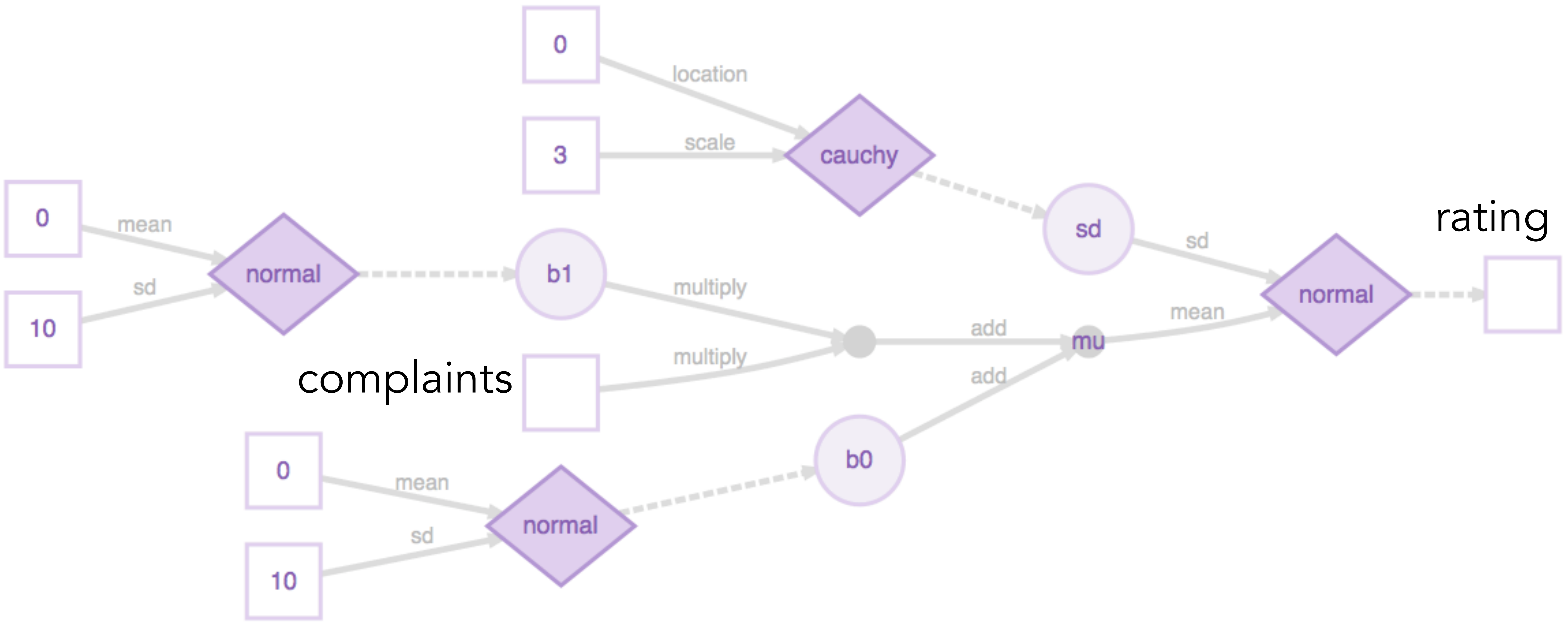


Truncated Cauchy prior on the standard deviation

weakly informative priors (allow for a wide range of possible values)

Graphical representation of the model

```r
# plotting
plot(m)
```



Inference via sampling

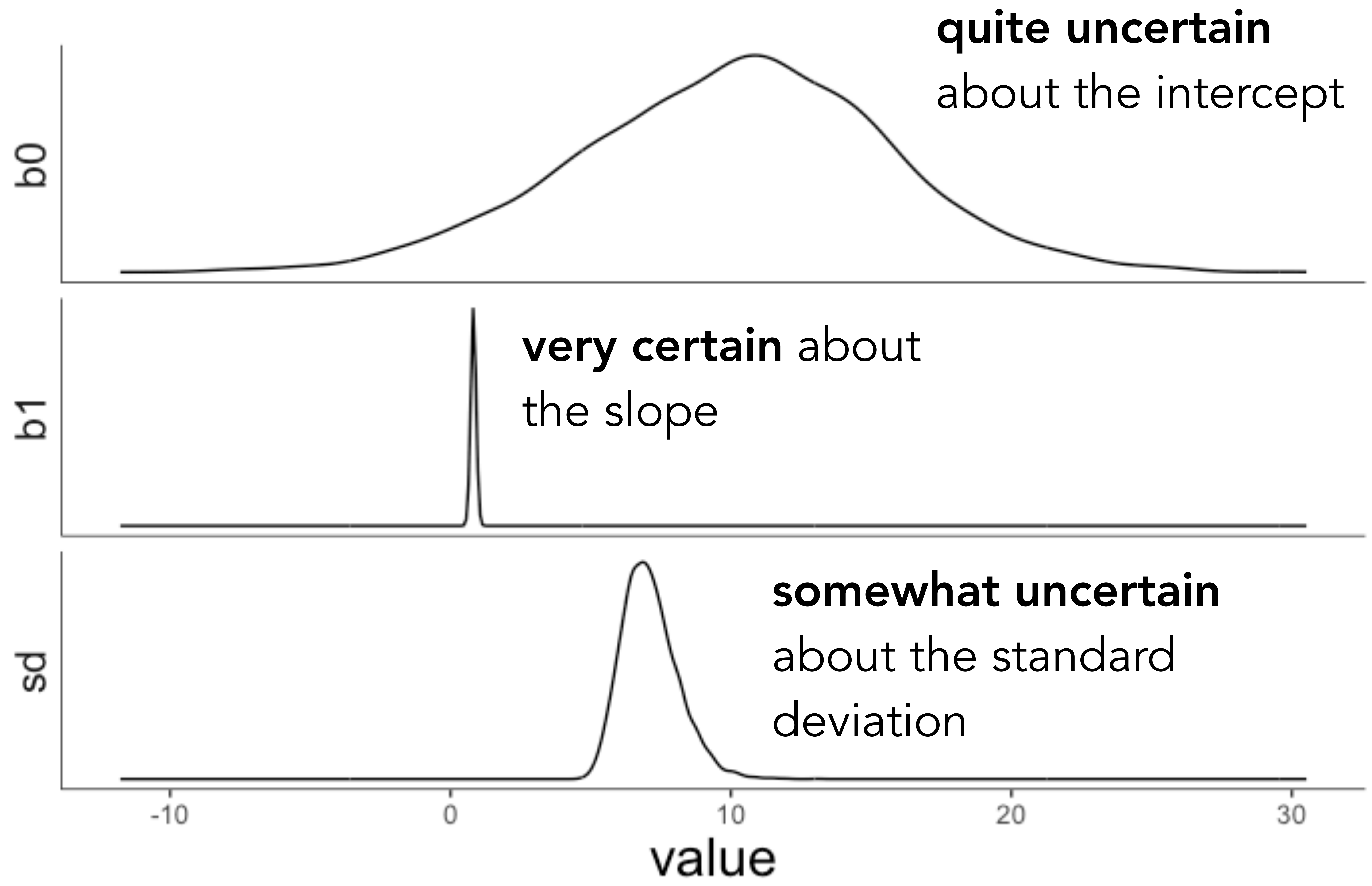
```r
# sampling
draws = mcmc(m, n_samples = 1000)

# tidy up the draws
df.draws = tidy_draws(draws) %>%
  clean_names()
```

Markov Chain Monte Carlo inference

chain	iteration	draw	b0	b1	sd
1	1	1	6.08	0.87	7.60
1	2	2	1.12	0.95	7.66
1	3	3	-1.83	0.99	7.01
1	4	4	-4.23	1.02	6.64
1	5	5	3.26	0.87	7.96
1	6	6	-1.04	0.98	7.67
1	7	7	-0.83	0.97	10.12
1	8	8	-1.41	0.97	8.02
1	9	9	9.46	0.81	6.30
1	10	10	10.02	0.84	6.57

each of these is a solution for explaining the data

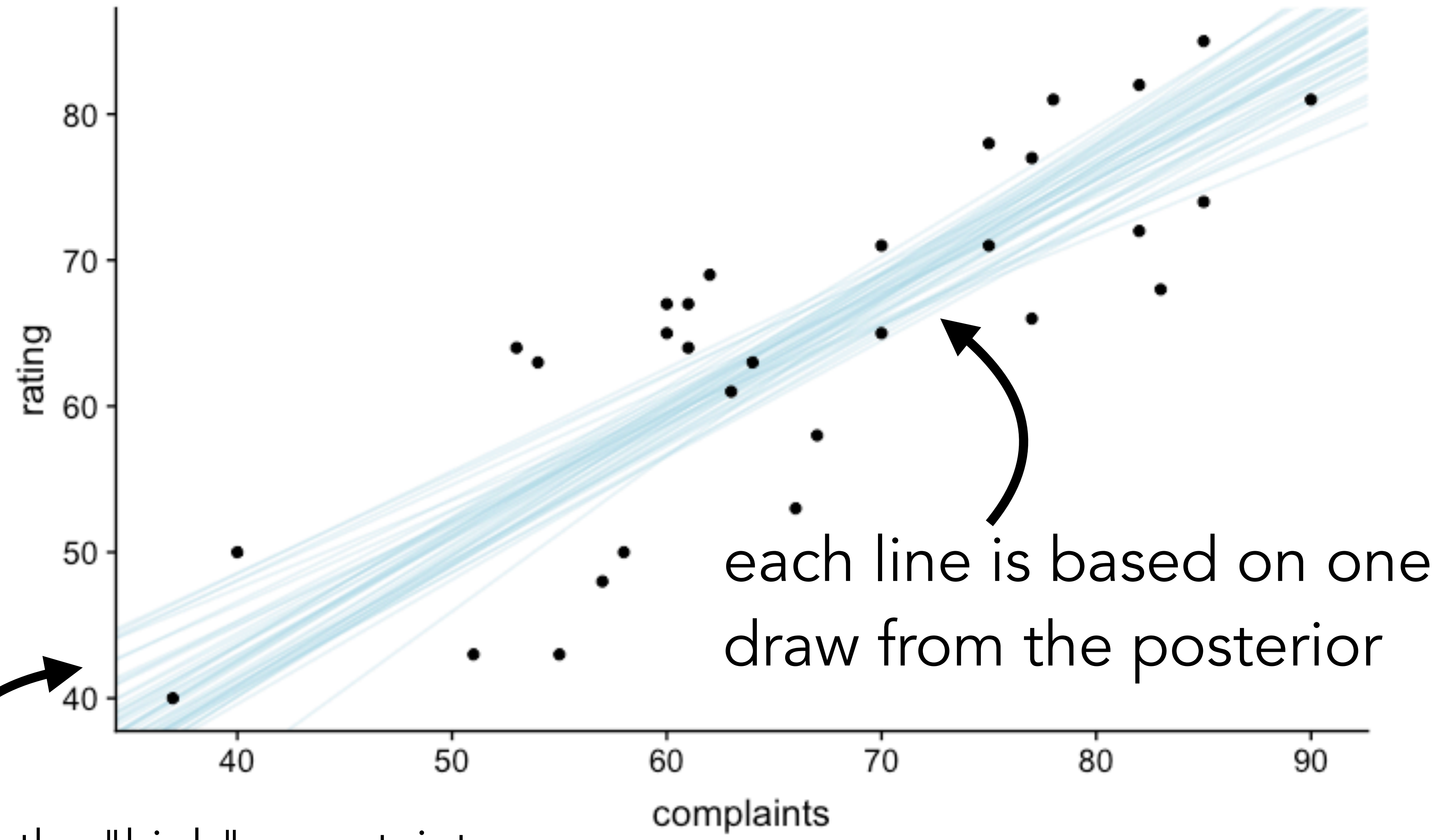
Visualize the posterior



Posterior distribution over the three parameters in the model

Visualize the model predictions

```r
ggplot(data = df.attitude,
       mapping = aes(x = complaints,
                     y = rating)) +
  geom_abline(data = df.draws %>%
                sample_n(size = 50),
              aes(intercept = b0,
                  slope = b1),
              alpha = 0.3,
              color = "lightblue") +
  geom_point()
```

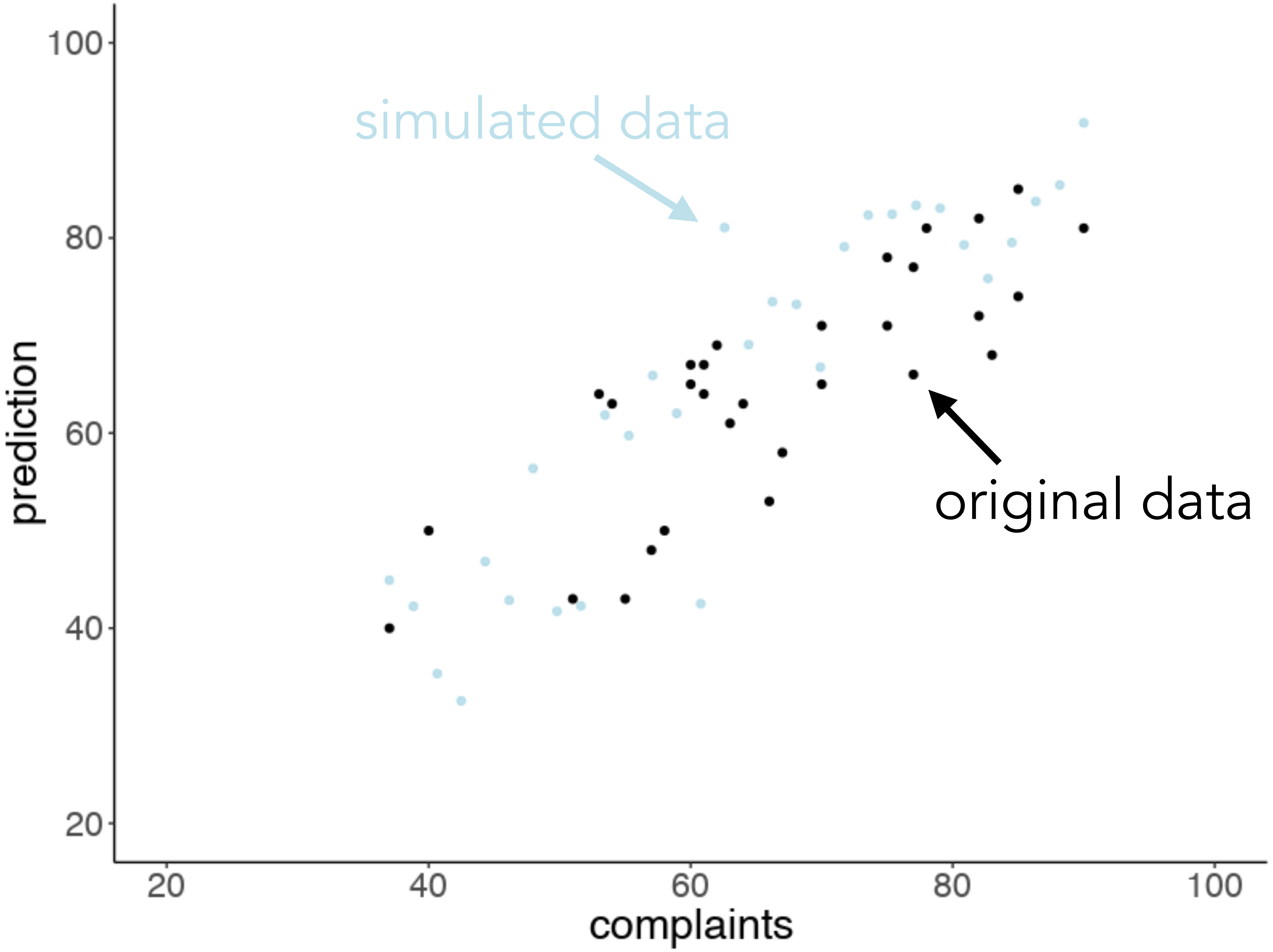


explains the "high" uncertainty about the intercept

chain	iteration	draw	b0	b1	sd
1	1	1	6.08	0.87	7.60
1	2	2	1.12	0.95	7.66
1	3	3	-1.83	0.99	7.01
1	4	4	-4.23	1.02	6.64
1	5	5	3.26	0.87	7.96
1	6	6	-1.04	0.98	7.67
1	7	7	-0.83	0.97	10.12
1	8	8	-1.41	0.97	8.02
1	9	9	9.46	0.81	6.30
1	10	10	10.02	0.84	6.57

Posterior predictive check

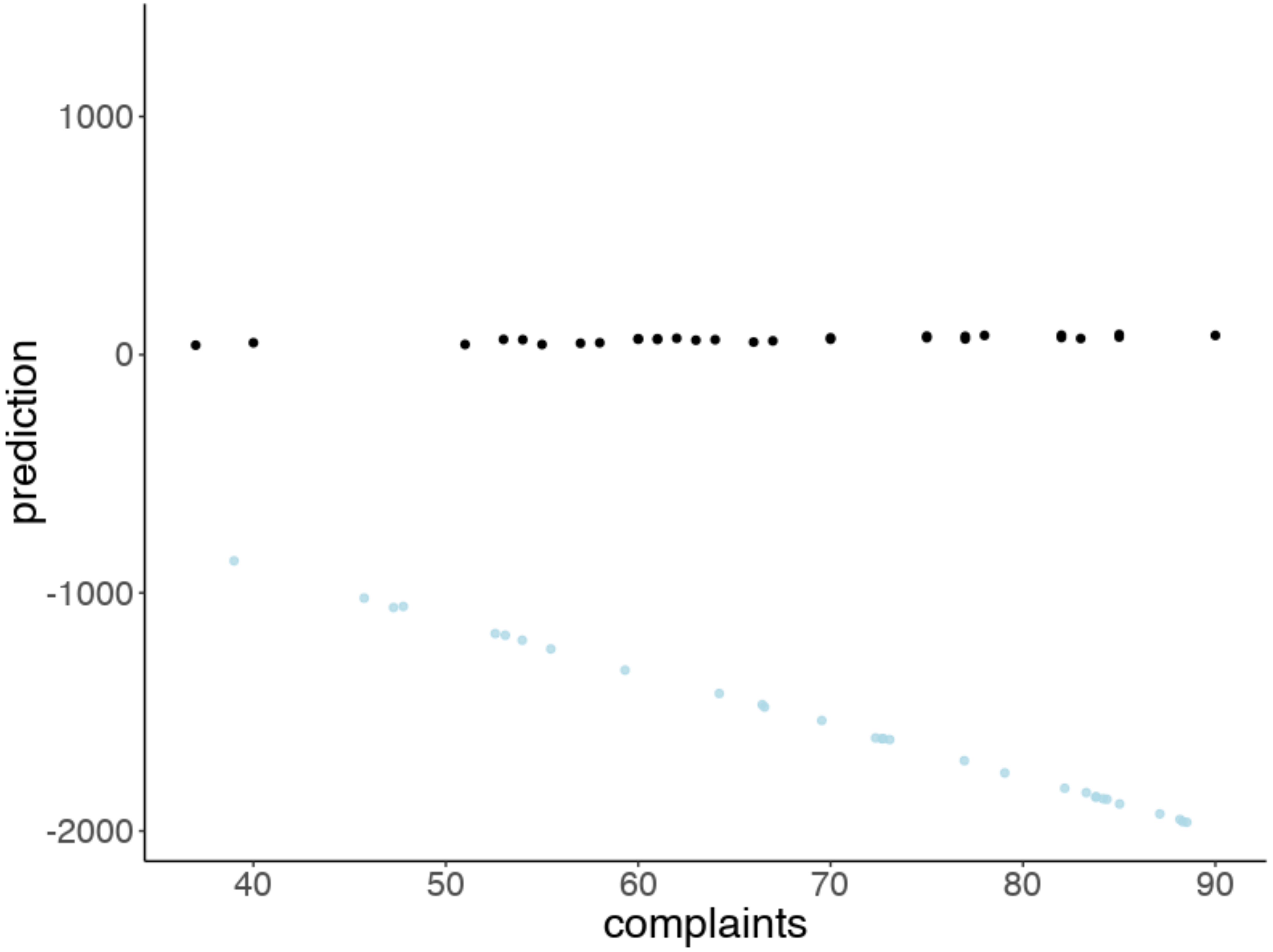
1. sample parameters from the **posterior distribution**
2. generate data using these parameters (using the likelihood function)



Hypothetical outcome plot

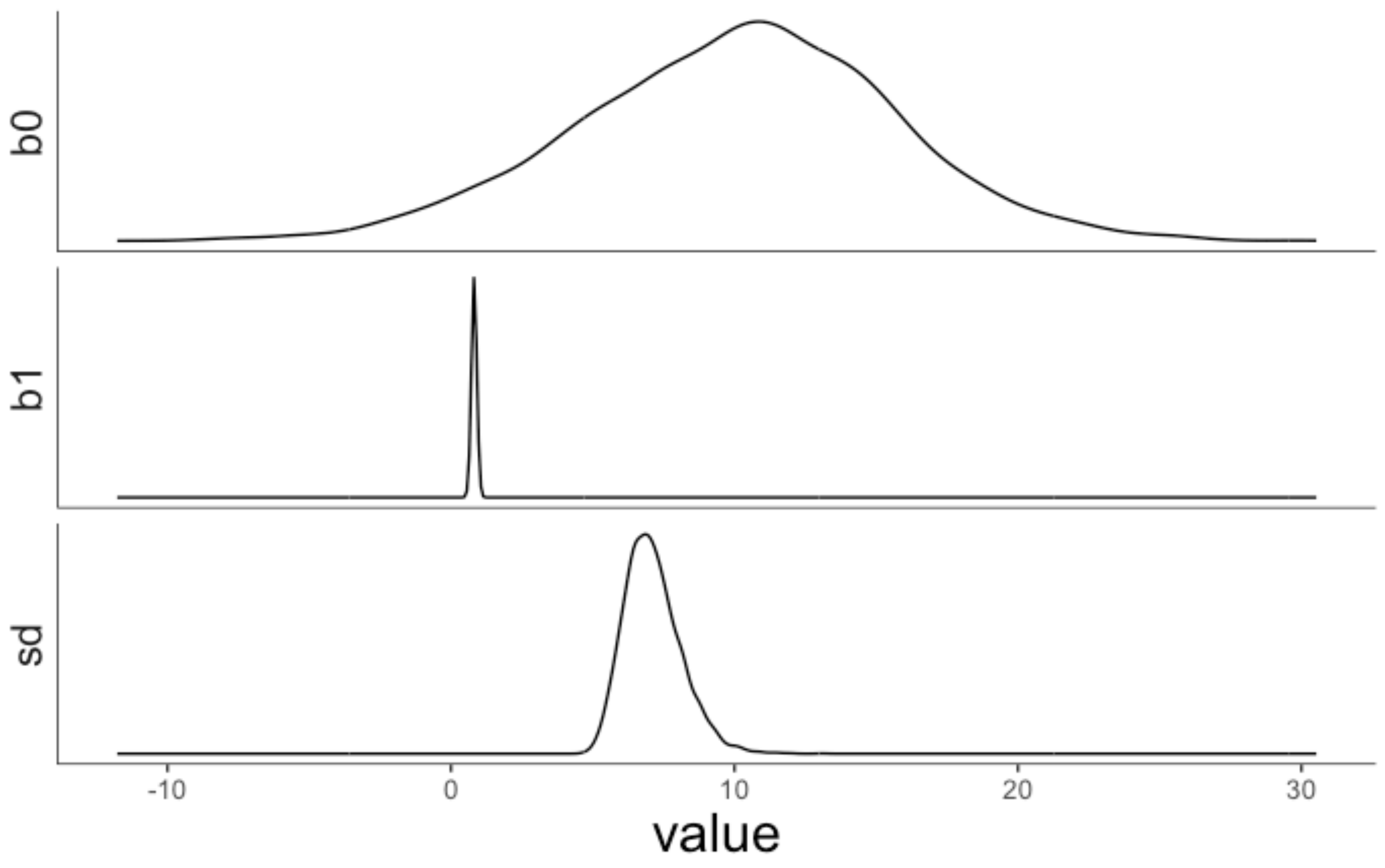
Prior predictive check

1. sample parameters from the **prior distribution**
2. generate data using these parameters (using the likelihood function)



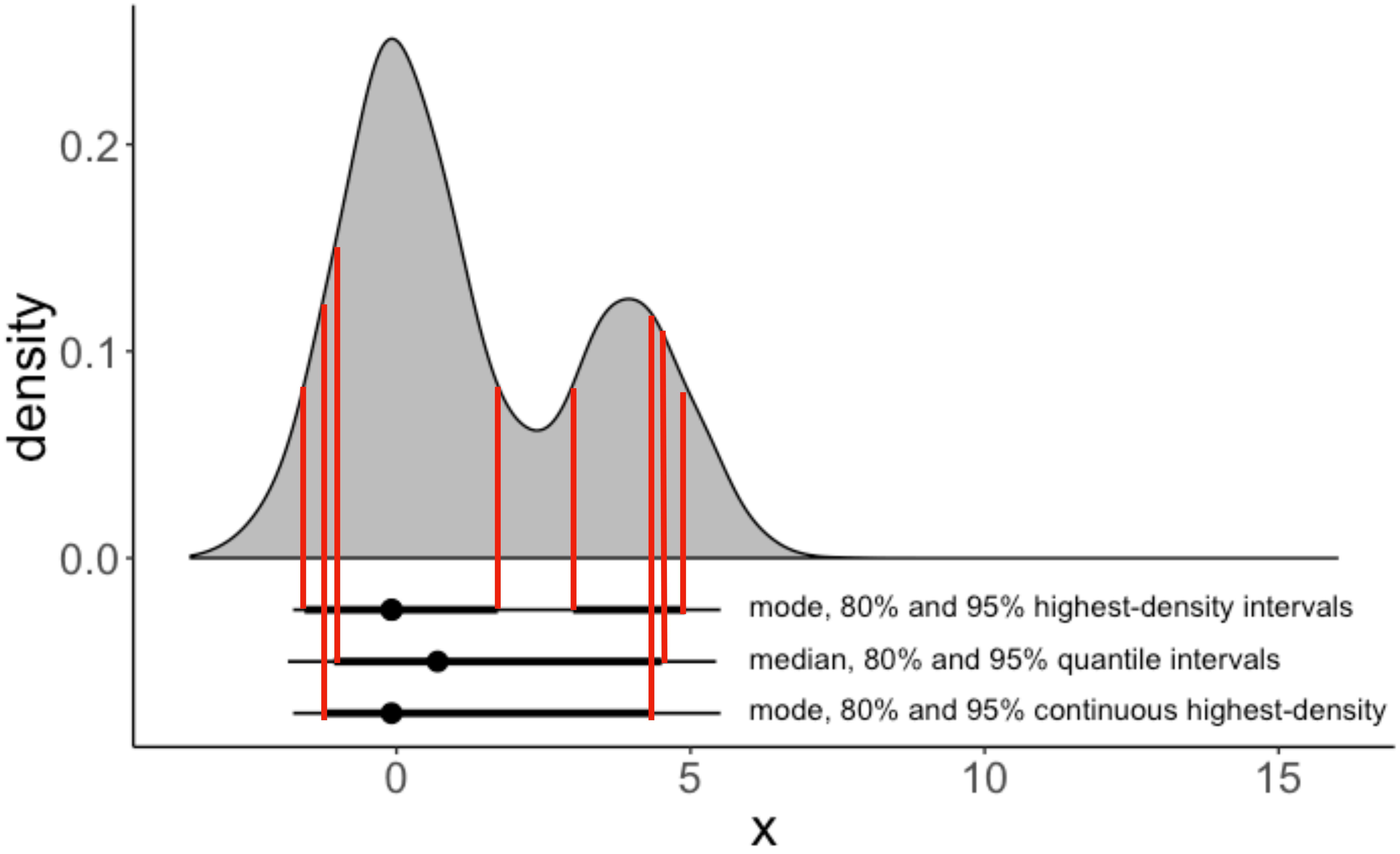
Hypothetical outcome plot

Summarizing results



- Posterior over each parameter is the result of the Bayesian data analysis.
- no p-values
- no confidence intervals

Different kinds of credible intervals



Summary

- Quick recap
- Causality
  - Patterns of inference
  - Should I control?
  - Mediation
  - Moderation
- Bayesian data analysis
  - Comparison between frequentist and Bayesian data analysis
  - Quick flash from the past
  - Flipping coins
  - What affects the posterior?
  - Ingredients: likelihood, prior, inference
  - Doing Bayesian data analysis

Feedback

How was the pace of today's class?

How happy were you with today's class overall?

What did you like about today's class? What could be improved next time?

Thank you!