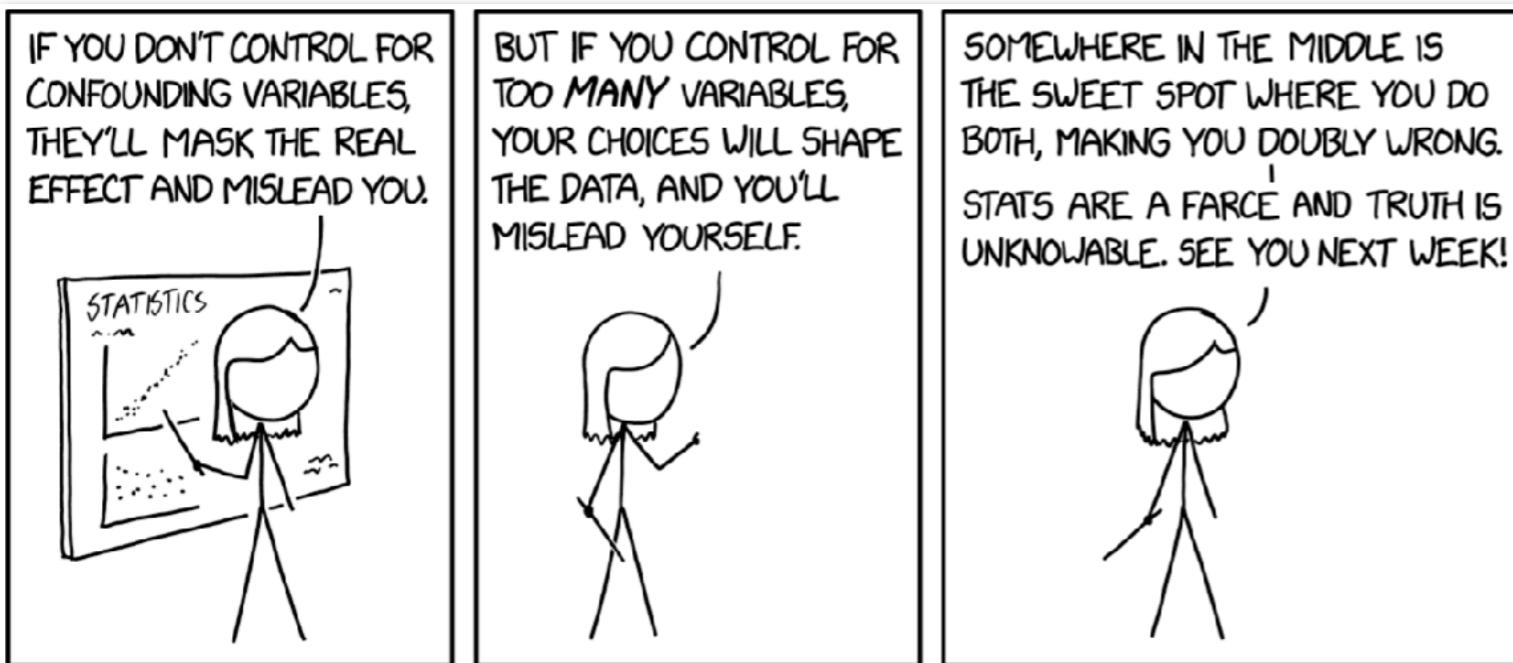


Causation



Chat

Would you rather travel back in time to meet your ancestors or would you rather go to the future to meet your descendants?

To: Everyone ▾ More ▾

Type message here...

COLLABORATIVE PLAYLIST

psych252

<https://tinyurl.com/psych252spotify22>

PLAY ...

02/11/2022

Logistics

Project proposal

Project proposal

IT'S TIME FOR A...

GROUP ASSIGNMENT!!

**Members of teams
will all the get
same grade!**

**maximum 3 team
members**



Didn't attend
any group
meetings



Doesn't
understand
the material



Gave the
presentation
but obviously
didn't know
what he was
even saying



Who is
this guy



"You can
use my
printer"



Did all the
research, wrote
paper, composed
presentation

Project proposal

The screenshot shows a GitHub Classroom interface. At the top, there's a dark header bar with the GitHub Classroom logo on the left and navigation links like "GitHub Education", a search icon, a profile icon, and a help icon on the right. Below the header, the main content area has a light gray background. It displays a repository card for "psych252". The repository name "psych252" is in a small, gray font. Below it, the title "Accept the assignment – final-project" is displayed, where "final-project" is in green. A descriptive text follows: "Once you accept this assignment, you will be granted access to the `final-project-tobiasgerstenberg` repository in the [psych252](#) organization on GitHub." At the bottom of the card is a green button labeled "Accept this assignment".



if you work as a team, have one person accept the assignment and the others can then join that repository

<https://classroom.github.com/a/g3qRXbG>

Project proposal

The screenshot shows a GitHub repository page for 'psych252 / final-projects'. The repository is private, has 3 watches, 0 stars, and 0 forks. It contains 5 commits, 1 branch, 0 releases, and 1 contributor (tobiasgerstenberg). The latest commit was 76f3633, 14 minutes ago. The repository structure includes 'code/R', 'data', 'figures', 'papers', 'presentation', 'writeup', '.gitignore', and 'README.md'. Below the repository details, there's a 'Final project' section with 'Starter code for your final project.' and a 'General points' section containing a bulleted list of guidelines for folder and file naming, relative paths, organization, and a note about empty folders. At the bottom, there's a 'Repository structure' section.

Final project

Starter code for your final project.

General points

- for folder and file names:
 - don't use white space in either folder or filenames, use an underscore "_" instead
 - (almost always) use lower case only
- always use relative paths in your code
 - for example, to save a figure from an R script inside the `code/R/` folder the path should be `"../../figures/figure_name.pdf"`
- keep your folder structure organized
 - we recommend adhering to the folder structure in this repository
 - more complex projects may have additional folders such as `videos/`, `papers/`, ...
- note: some of the folders are empty except for a `.keep` file
 - the `.keep` file is just there to make sure that github includes the otherwise empty folder
 - feel free to delete the `.keep` file once you've added another file to that folder

Repository structure

- each team will have their own private github repository
- all work on your final project should happen within this repository
- you can get **github** help in section
- post on Ed Discussion in case you experience any problems getting set up (use **final_project** tag)

Project proposal

RMarkdown template



The screenshot shows the RStudio interface with the following details:

- Project Area:** Shows a file named "project_proposal.Rmd".
- Code Editor:** Displays the RMarkdown code for a project proposal. The code includes metadata like title, subtitle, author, date, and output format, followed by sections for instructions and research questions.
- Environment:** Shows the global environment is empty.
- Files:** Shows the project structure: final-projects > code > R. It lists files: project_proposal.Rmd, project_report.html, project_report.Rmd, psych252-final-project.Rproj, and references.
- Console:** Shows the R version (4.1.2) and the current working directory (~/Documents/work/projects/psych252/final-projects/code/R).

Upload the
pdf to canvas

Project proposal

The project proposal is due on
Thursday, February 17th at 8pm

Project proposal

W22-PSYCH-252-01 > Files > final_project > proposal > 2021

Search for files Q 0 items selected + Folder Upload ⋮

Name	Date Created	Date Modified	Modified By	Size	⋮
anjie_cao.pdf	2:59pm	2:59pm	Tobias Gerste...	236 KB	<input checked="" type="checkbox"/>
danyang_fan.pdf	2:59pm	2:59pm	Tobias Gerste...	259 KB	<input checked="" type="checkbox"/>
hannah_marshall.pdf	2:59pm	2:59pm	Tobias Gerste...	238 KB	<input checked="" type="checkbox"/>
madi_jamie_catie.pdf	2:59pm	2:59pm	Tobias Gerste...	247 KB	<input checked="" type="checkbox"/>

8% of 5.2 GB used All My Files



examples of project proposals from prior years

Plan for today

- Quick recap
- Model comparison
 - Cross-validation in action
 - AIC and BIC
- Causation vs. correlation
- Controlling for variables
- Mediation
- Moderation

Quick recap

Quick recap: Power analysis via simulation

Power simulation recipe

- assume:
 - α , n , effect size
- simulate a large number of data sets of size n with the specified effect size
- for each data set, run a statistical test to calculate the p-value
- determine the probability of rejecting the H_0 (given that H_1 is true)

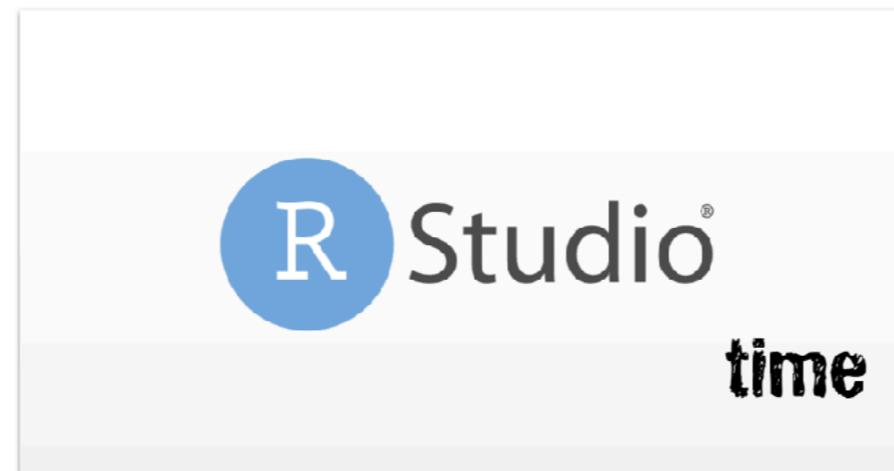
Let's simulate ...

```
1 # make reproducible
2 set.seed(1)
3
4 # number of simulations
5 n_simulations = 5
6
7 # run simulation
8 expand_grid(n = seq(10, 40, 2),
9   simulation = 1:n_simulations,
10   p = 0.75) %>%
11   mutate(index = 1:n(),
12     .before = n) %>%
13   group_by(index, n, p, simulation) %>%
14   mutate(response = rbinom(n = 1,
15     size = n,
16     prob = p),
17     p.value = binom.test(x = response,
18     n = n,
19     p = 0.5,
20     alternative = "two.sided")$p.value) %>%
21   group_by(n, p) %>%
22   summarize(power = sum(p.value < 0.05) / n())
```

n	p	power
10	0.75	0.2
12	0.75	0.2
14	0.75	0.4
16	0.75	0.2
18	0.75	0.6
20	0.75	0.8
22	0.75	0.6
24	0.75	0.4
26	0.75	0.6
28	0.75	0.8
30	0.75	0.8

13

18



map()
unnest()

21

12

Quick recap: Model comparison

The general procedure

1. Define H_0 as Model C (compact) and H_1 as Model A (augmented)
2. Fit model parameters to the data
3. Calculate the proportional reduction of error (PRE) in our sample
4. Decide whether the augmented model is **worth it** by comparing the observed PRE in our sample to the sampling distribution of PRE (assuming that H_0 is true)

Any problems with our approach?

sometimes it doesn't work ...

Model C

$$\text{balance}_i = \beta_0 + \epsilon_i$$

$$\text{balance}_i = \beta_0 + \beta_1 \cdot \text{student}_i + \epsilon_i$$

$$\text{balance}_i = \beta_0 + \beta_1 \cdot \text{student}_i + \epsilon_i$$

$$\text{balance}_i = \beta_0 + \beta_1 \cdot \text{student}_i + \epsilon_i$$

Model A

$$\text{balance}_i = \beta_0 + \beta_1 \cdot \text{student}_i + \beta_2 \cdot \text{age}_i + \epsilon_i$$

$$\text{balance}_i = \beta_0 + \beta_1 \cdot \text{student}_i + \beta_2 \cdot \text{age}_i + \epsilon_i$$

$$\text{balance}_i = \beta_0 + \beta_1 \cdot \text{age}_i + \epsilon_i$$

$$\text{balance}_i = \beta_0 + \beta_1 \cdot \text{age}_i + \beta_2 \cdot \text{degree}_i + \epsilon_i$$

23

24

Tools for model comparison

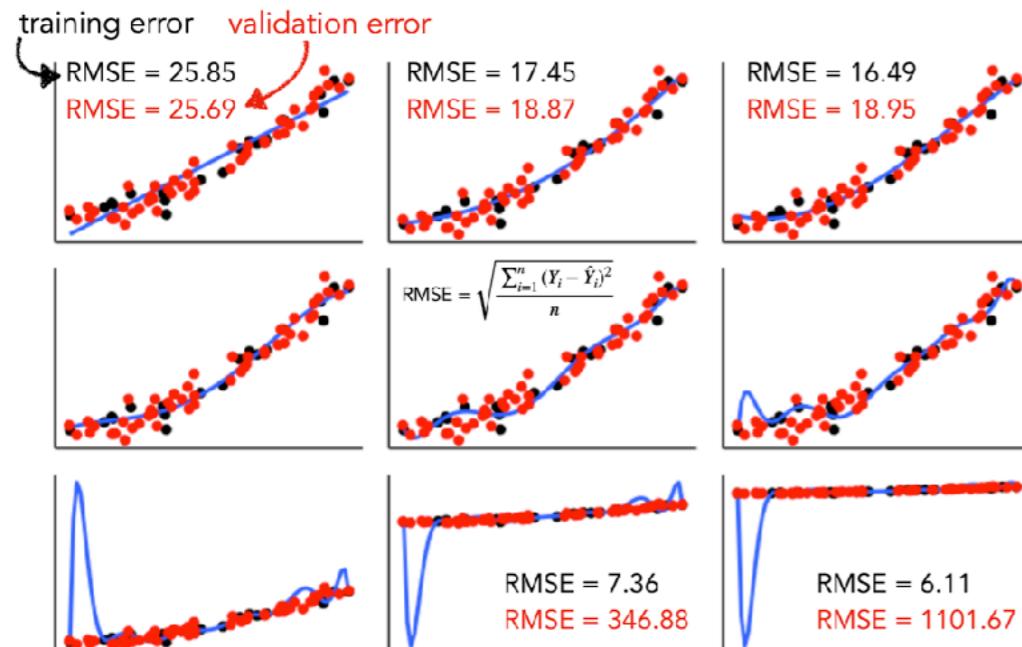
- **anova()** : compare a compact model with an augmented model via the F-test
 - problem: only works for nested models (where the augmented model contains all the predictors of the compact model and more)
- **What if we want to compare models that aren't nested?**
 - Cross-validation
 - AIC and BIC
 - Bayesian data analysis (we'll get there soon)

25

13

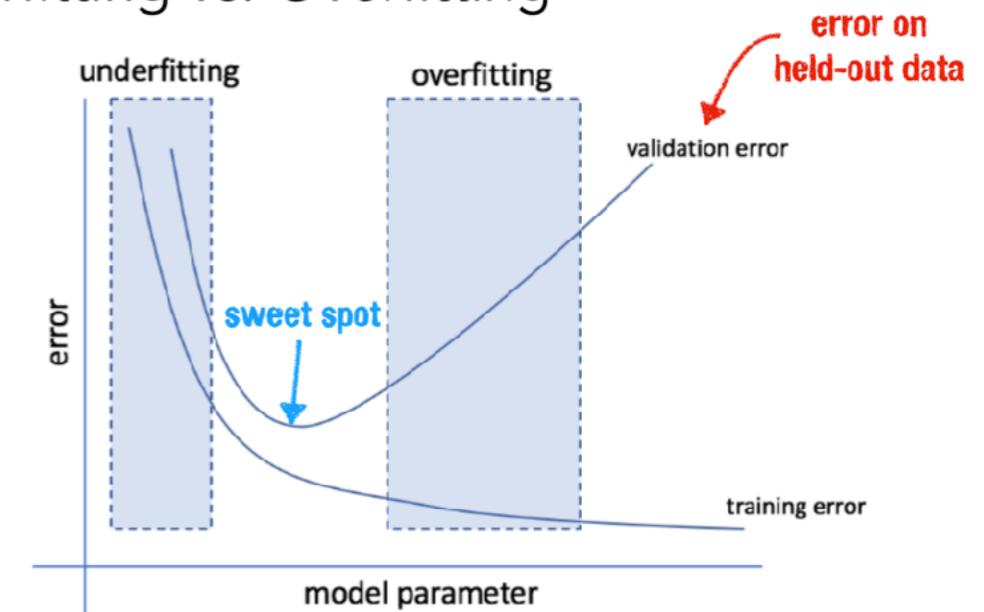
Quick recap: Cross-validation

Which model describes the data best?



28

Underfitting vs. Overfitting



the goal is to find the **sweet spot** between underfitting and overfitting

31

k-fold crossvalidation



33

	Full data set				
Iteration 1	Test	Train	Train	Train	Train
Iteration 2	Train	Test	Train	Train	Train
Iteration 3	Train	Train	Test	Train	Train
Iteration 4	Train	Train	Train	Test	Train
Iteration 5	Train	Train	Train	Train	Test

Monte Carlo crossvalidation

random splits into training and test data
crossv_mc ($n = 50$, $test = 0.5$)
 number of training-test splits
 proportion of test data in each split

48

14

Model comparison

Cross-validation in action



Studio[®]

time

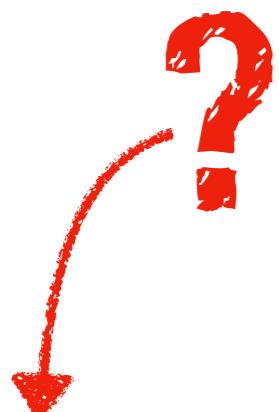
AIC and BIC

AIC and BIC

- AIC = Akaike Information Criterion
- BIC = Bayesian Information Criterion

not that much Bayesian about it ...

$$\text{AIC} = 2k - 2 \ln(\hat{L})$$



$$\text{BIC} = \ln(n)k - 2 \ln(\hat{L})$$

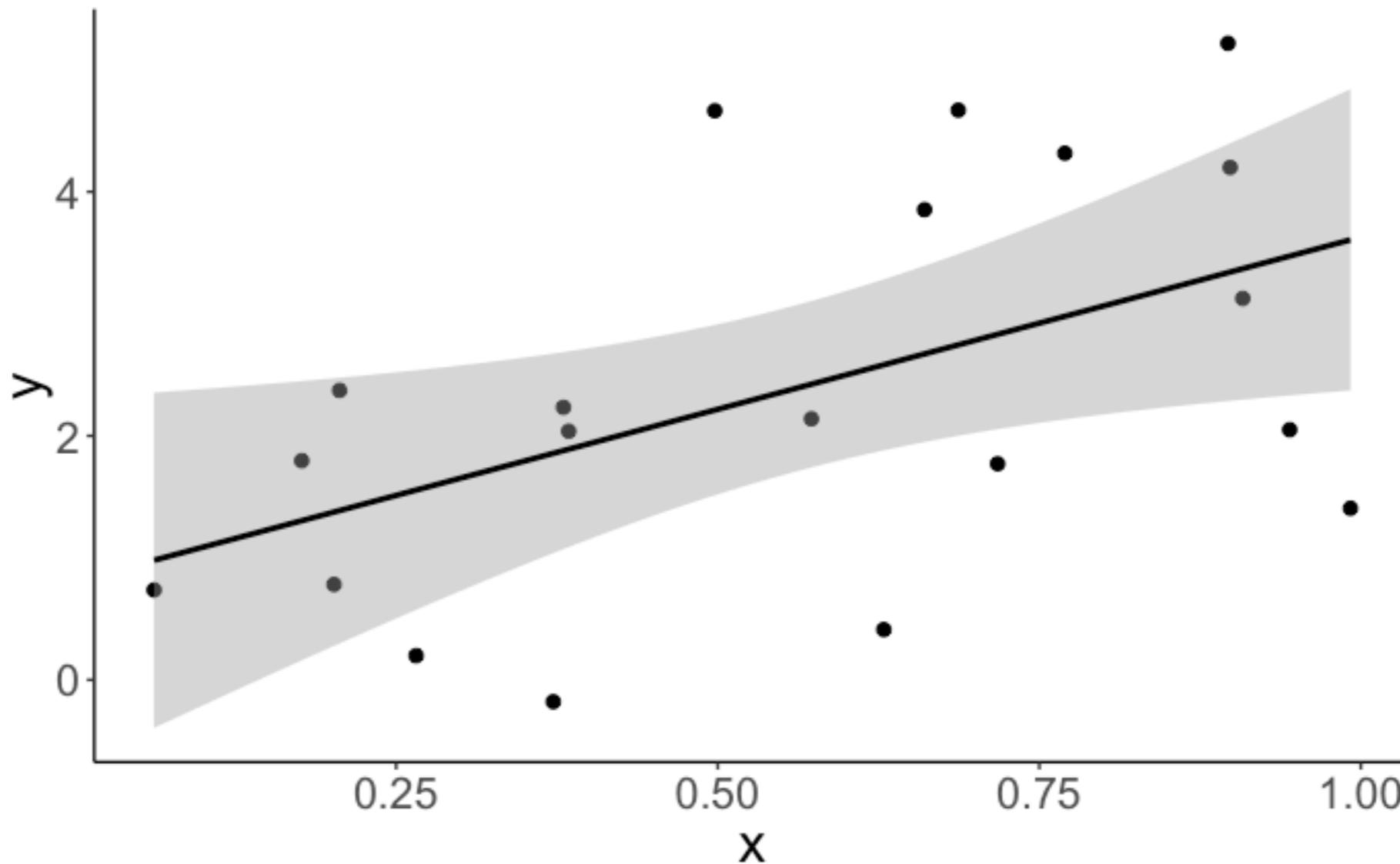
- \hat{L} = maximized value of the likelihood function of the model
 k = number of parameters in the model
 n = number of observations

AIC and BIC

- How do we get the likelihood of our model?
 - in a linear regression, minimizing least squares is equivalent to maximizing the likelihood of the data given the model
- Assumptions of the linear model:
 - residuals are normally distributed with:
 - mean = 0 and sd = sigma
 - calculate overall likelihood by computing the likelihood of each residual, and then multiplying

AIC and BIC

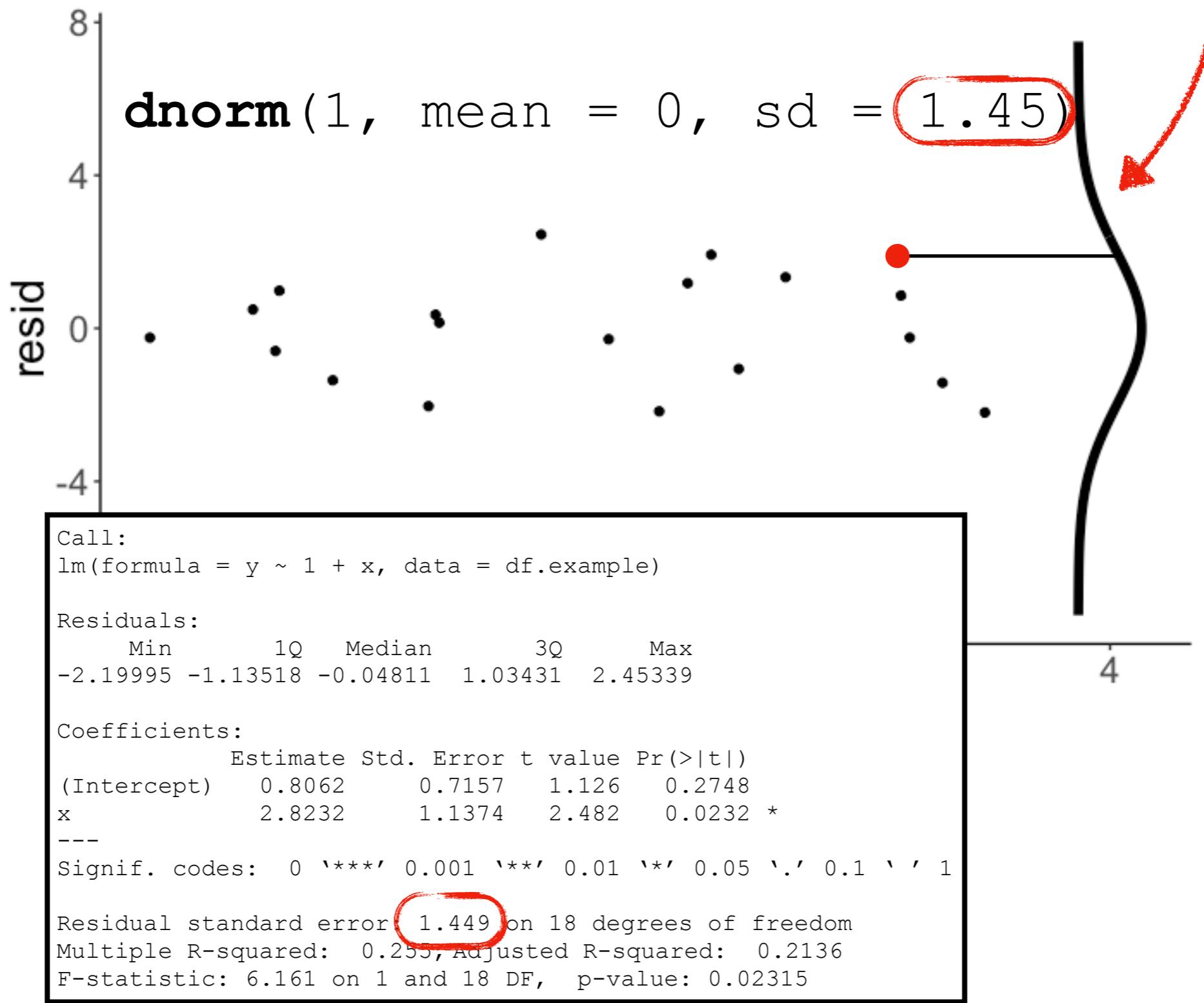
data with linear model fit



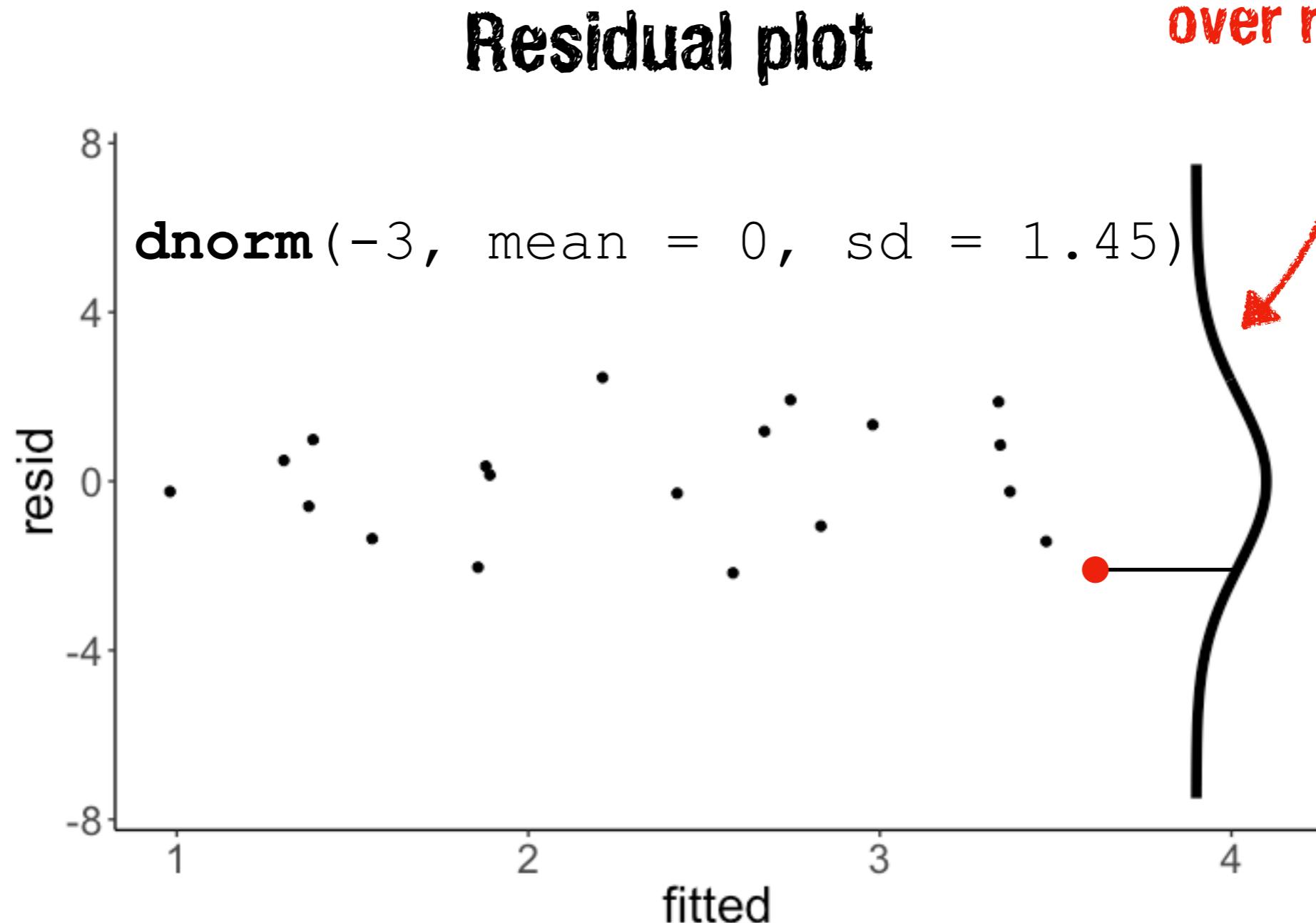
AIC and BIC

normal distribution
over residuals

Residual plot



AIC and BIC



normal distribution
over residuals

since the data points are independent, we can calculate the overall likelihood by multiplying the likelihood of each observation

AIC and BIC

```

1 # generate some data
2 df.like = tibble(
3   x = runif(20, min = 0, max = 1),
4   y = 1 + 3 * x + rnorm(20, sd = 2)
5 )
6
7 # fit the model
8 fit = lm(formula = y ~ x,
9           data = df.like)
10
11 # model summary
12 fit %>%
13   glance()

```

`dnorm(1.88, mean = 0, sd = 1.45) = 0.12`

x	y	fitted	resid	likelihood
0.90	5.22	3.34	1.88	0.12
0.27	0.20	1.56	-1.36	0.18
0.37	-0.18	1.86	-2.04	0.10
0.57	2.14	2.42	-0.28	0.27
0.91	3.13	3.37	-0.24	0.27
0.20	0.78	1.38	-0.59	0.25
0.90	4.20	3.34	0.86	0.23
0.94	2.05	3.47	-1.42	0.17
0.66	3.85	2.67	1.18	0.20
0.63	0.41	2.58	-2.17	0.09

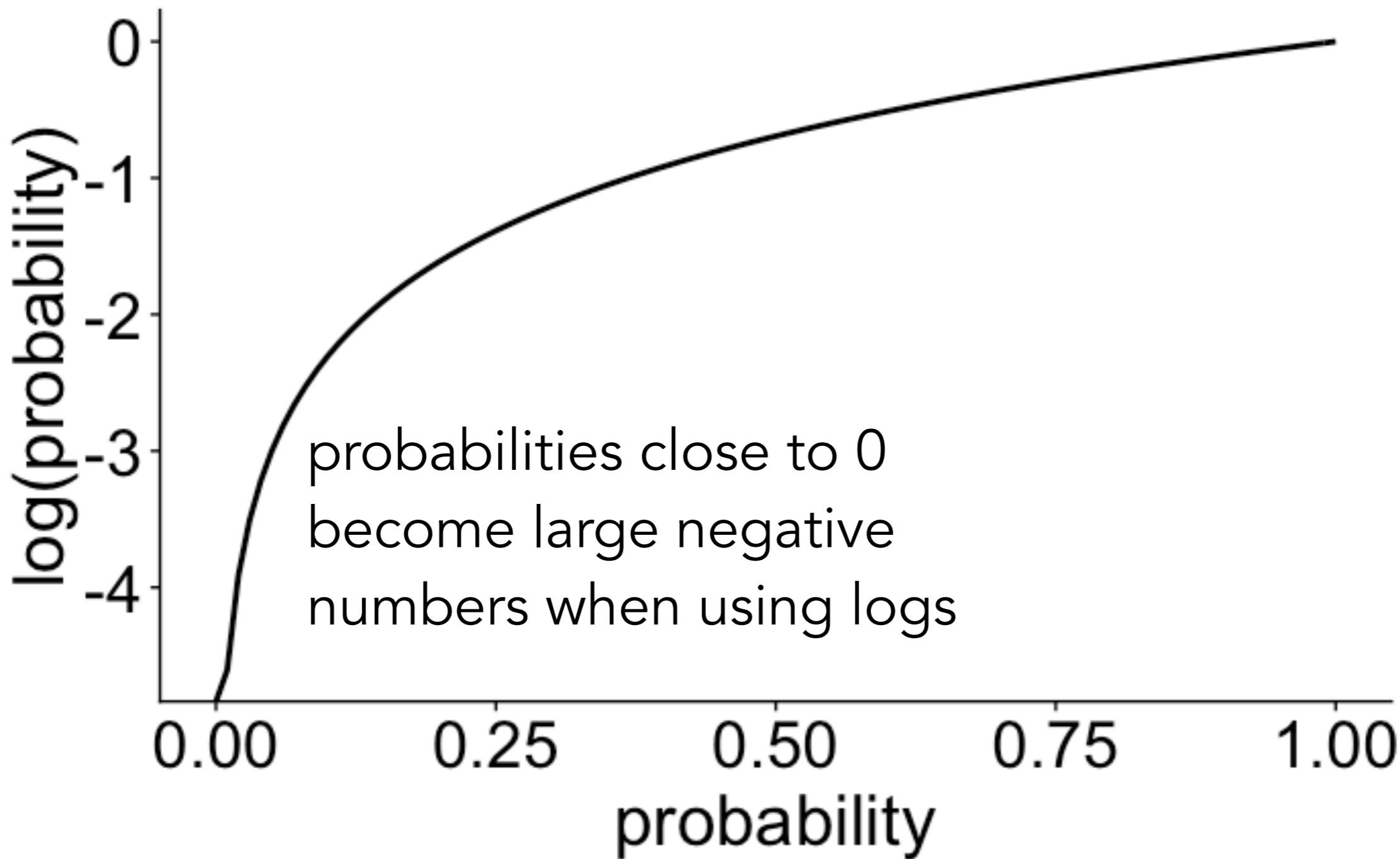
inferred standard
deviation of the error

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual
0.25	0.21	1.45	6.16	0.02	2	-34.74	75.47	78.46	37.77	18

$e \sim \mathcal{N}(\text{mean} = 0, \text{sd} = 1.45)$

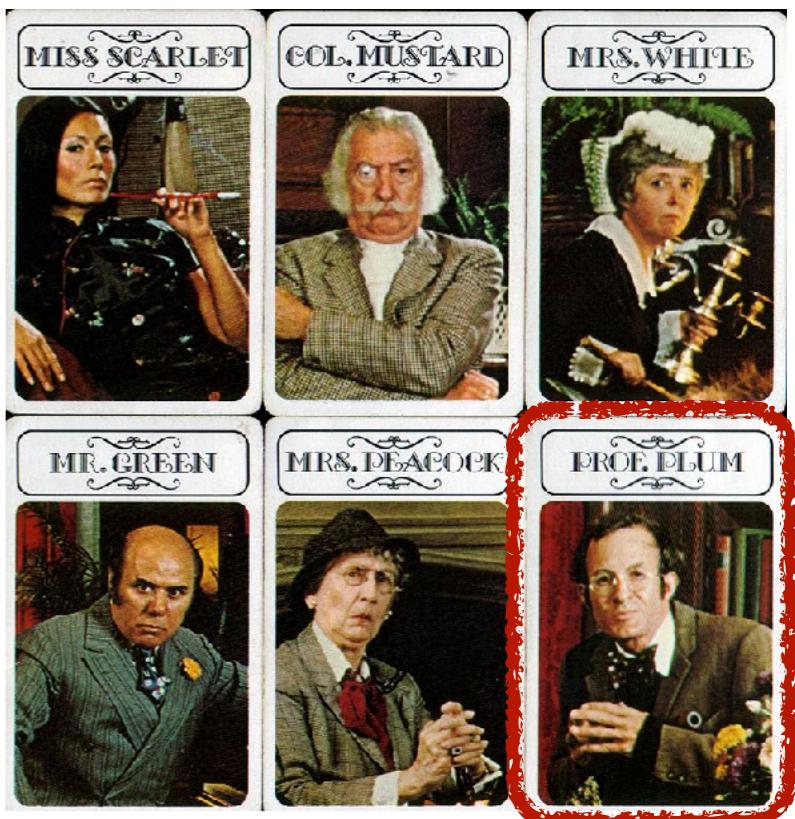
$$\sum_{i=1}^n \ln(\text{likelihood})$$

`log()` is your friend!



Clue guide to probability

Who?



- joint probability:

- if A and B are independent then

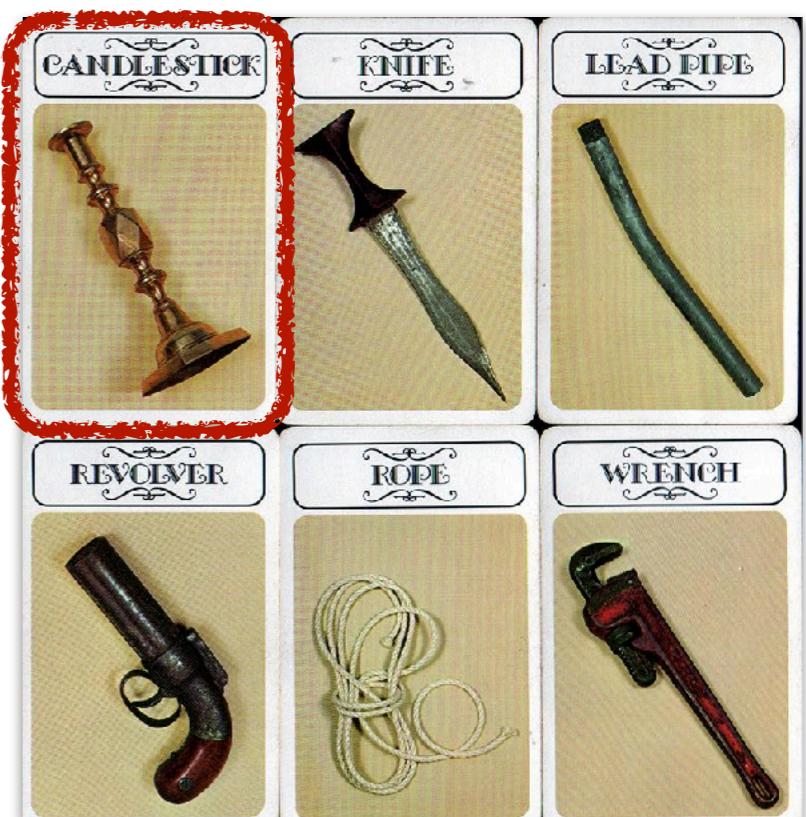
- Definition: $p(A, B) = p(A) \cdot p(B)$

- $p(\text{Prof Plum, candle stick}) =$

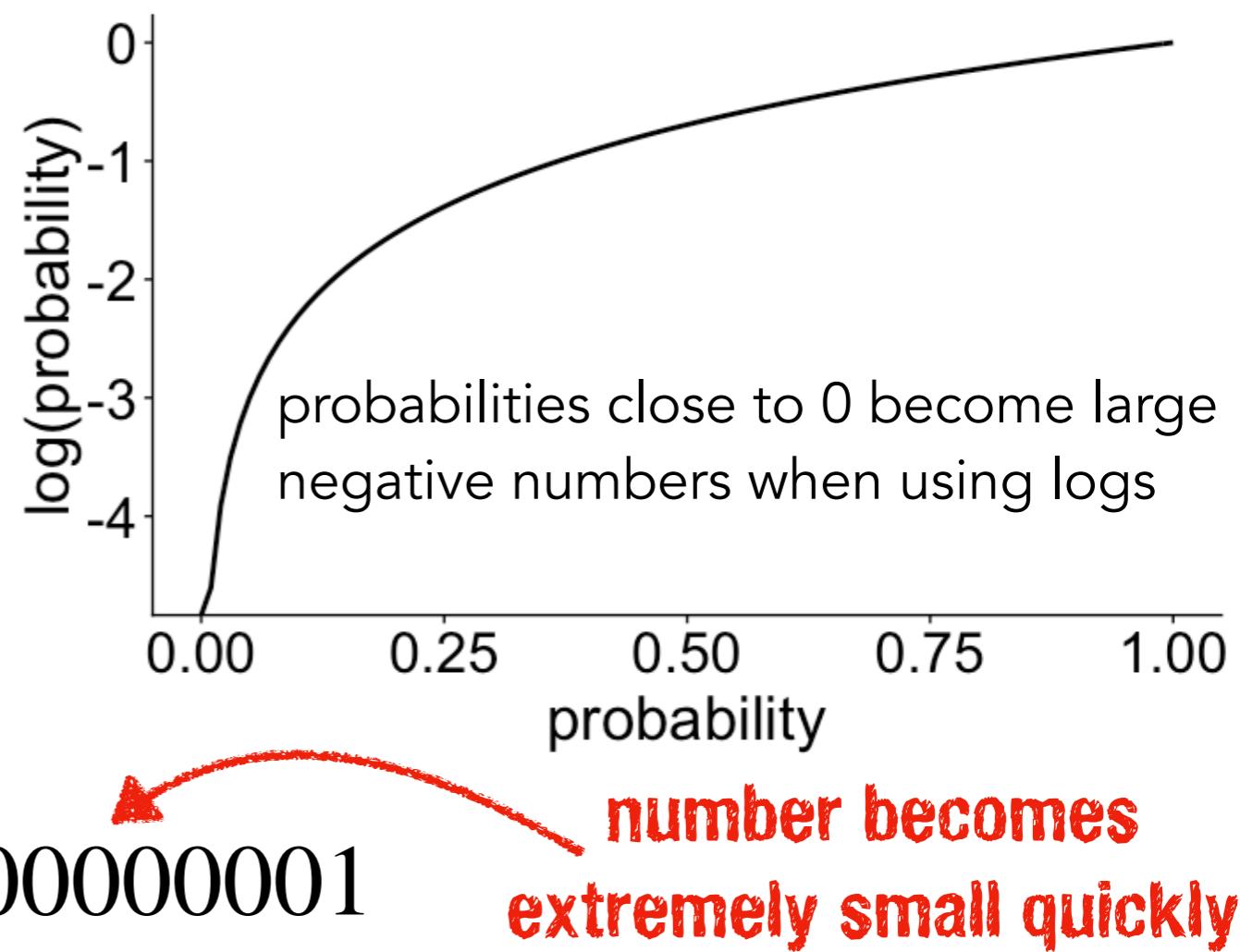
$$p(\text{Prof Plum}) \cdot p(\text{candle stick}) =$$

$$\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

What?



`log()` is your friend!



multiplying probabilities

$$0.01 \cdot 0.01 \cdot 0.01 \cdot 0.01 = 0.00000001$$

take `log()`

$$\log(0.01) = -4.60517$$

number becomes large
but that's ok

summing logs

$$(-4.60517) + (-4.60517) + (-4.60517) + (-4.60517) = -18.42068$$

transform back into probability

$$\exp(-18.42068) = 0.00000001$$

often not necessary since
we just use `logLikelihood`

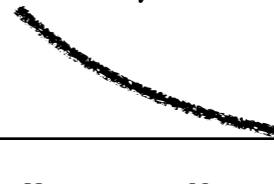
AIC and BIC

```

1 # generate some data
2 df.like = tibble(
3   x = runif(20, min = 0, max = 1),
4   y = 1 + 3 * x + rnorm(20, sd = 2)
5 )
6
7 # fit the model
8 fit = lm(formula = y ~ x,
9           data = df.like)
10
11 # model summary
12 fit %>%
13   glance()

```

`dnorm(1.88, mean = 0, sd = 1.45) = 0.12`



x	y	fitted	resid	likelihood
0.90	5.22	3.34	1.88	0.12
0.27	0.20	1.56	-1.36	0.18
0.37	-0.18	1.86	-2.04	0.10
0.57	2.14	2.42	-0.28	0.27
0.91	3.13	3.37	-0.24	0.27
0.20	0.78	1.38	-0.59	0.25
0.90	4.20	3.34	0.86	0.23
0.94	2.05	3.47	-1.42	0.17
0.66	3.85	2.67	1.18	0.20
0.63	0.41	2.58	-2.17	0.09

inferred standard
deviation of the error




r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual
0.25	0.21	1.45	6.16	0.02	2	-34.74	75.47	78.46	37.77	18

$e \sim \mathcal{N}(\text{mean} = 0, \text{sd} = 1.45)$

AIC and BIC

$$\text{AIC} = 2k - 2 \ln(\hat{L})$$

$$\text{BIC} = \ln(n)k - 2 \ln(\hat{L})$$

$\ln(\hat{L})$ = maximized value of the likelihood function of the model **-34.74**

k = number of parameters in the model **3**

n = number of observations **20**

the sd of the normal distribution modeling the residuals counts as a parameter

```
lm(formula = y ~ 1 + x, data = df.example)
```

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual
0.25	0.21	1.45	6.16	0.02	2	-34.74	75.47	78.46	37.77	18

AIC and BIC

$$\text{AIC} = 2k - 2 \ln(\hat{L}) = 2 \cdot 3 - 2 \cdot (-34.74) = 75.47$$

$$\text{BIC} = \ln(n)k - 2 \ln(\hat{L}) = \ln(20) \cdot 3 - 2 \cdot (-34.74) = 78.46$$

$\ln(\hat{L})$ = maximized value of the likelihood function of the model **-34.74**

k = number of parameters in the model **3**

n = number of observations **20**

the sd of the normal distribution modeling the residuals counts as a parameter

```
lm(formula = y ~ 1 + x, data = df.example)
```

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual
0.25	0.21	1.45	6.16	0.02	2	-34.74	75.47	78.46	37.77	18

AIC and BIC

$$\text{AIC} = 2k - 2 \ln(\hat{L})$$

$$\text{BIC} = \ln(n)k - 2 \ln(\hat{L})$$

- for both AIC and BIC, *lower* is better!
- neither provide a test of a model in the sense of testing a null hypothesis
 - AIC or BIC tell us nothing about the absolute quality of a model, only the quality relative to other models
- BIC generally penalizes free parameters more strongly than AIC (though it depends on the size of n)

ΔBIC	Evidence against higher BIC
0 to 2	Not worth more than a bare mention
2 to 6	Positive
6 to 10	Strong
>10	Very Strong

What shall I use when?

- Use it all!
- ideally, the different measures provide converging evidence

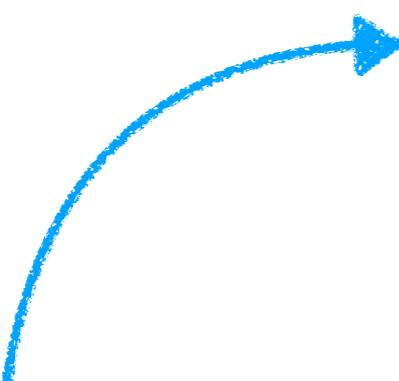
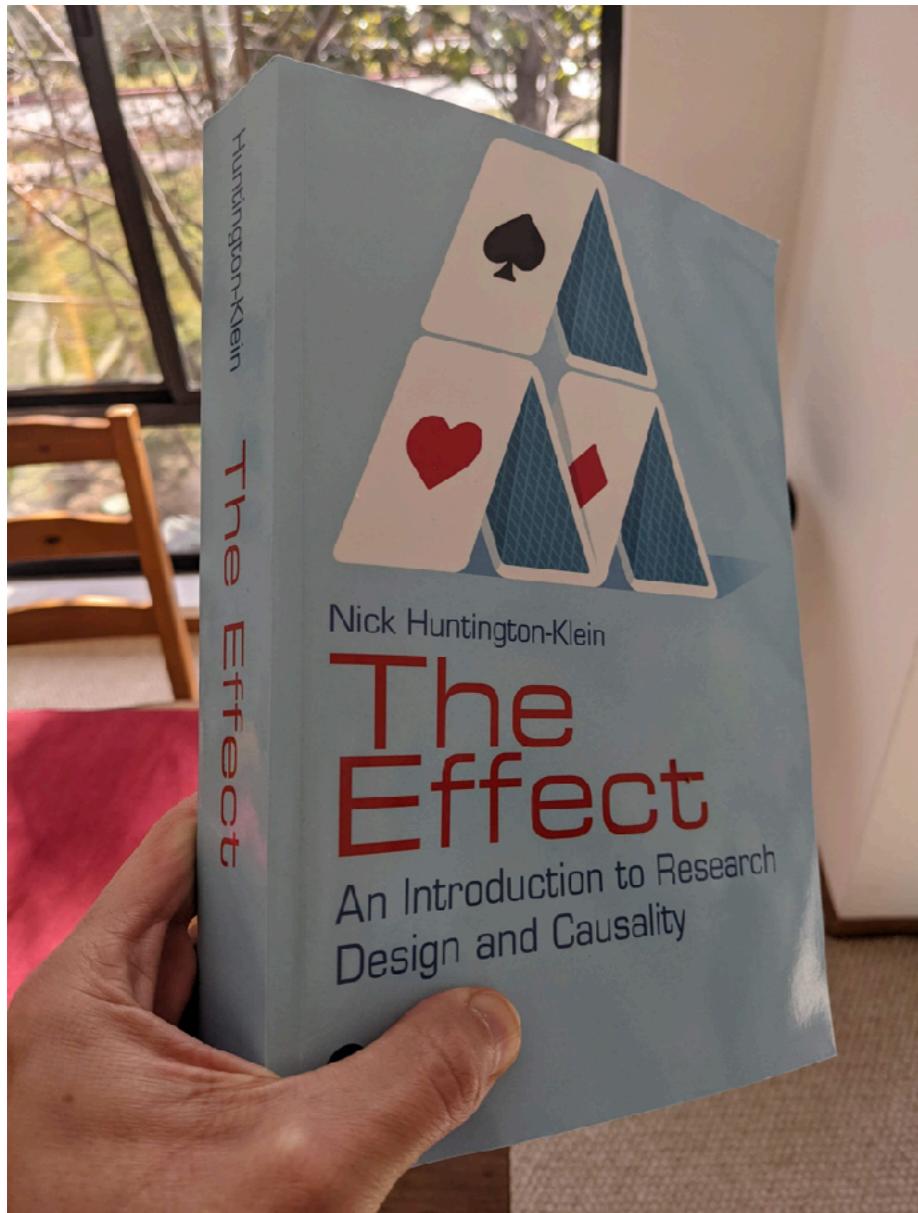
Table 2

Summary of the model results. Values for r and RMSE indicate means (with 5% and 95% quantiles in parentheses) based on 100 split-half cross-validation runs. BIC scores are based on running the models on the full data set.

Model	r	RMSE	BIC
Difference & pivotality	.86 (.66, .95)	10.56 (6.17, 17.21)	158.59
Difference	.70 (.30, .90)	26.92 (16.4, 40.6)	209.74
Pivotality	.63 (.41, .77)	14.23 (11.39, 17.54)	199.53
Optimality	.66 (.42, .84)	14.55 (10.54, 17.91)	199.47

Note: BIC = Bayesian Information Criterion (lower values indicate better model performance).

Causation vs. correlation



The Effect: An Introduction to Research Design and Causality

Search Additional Materials Revision and Updates

Introduction

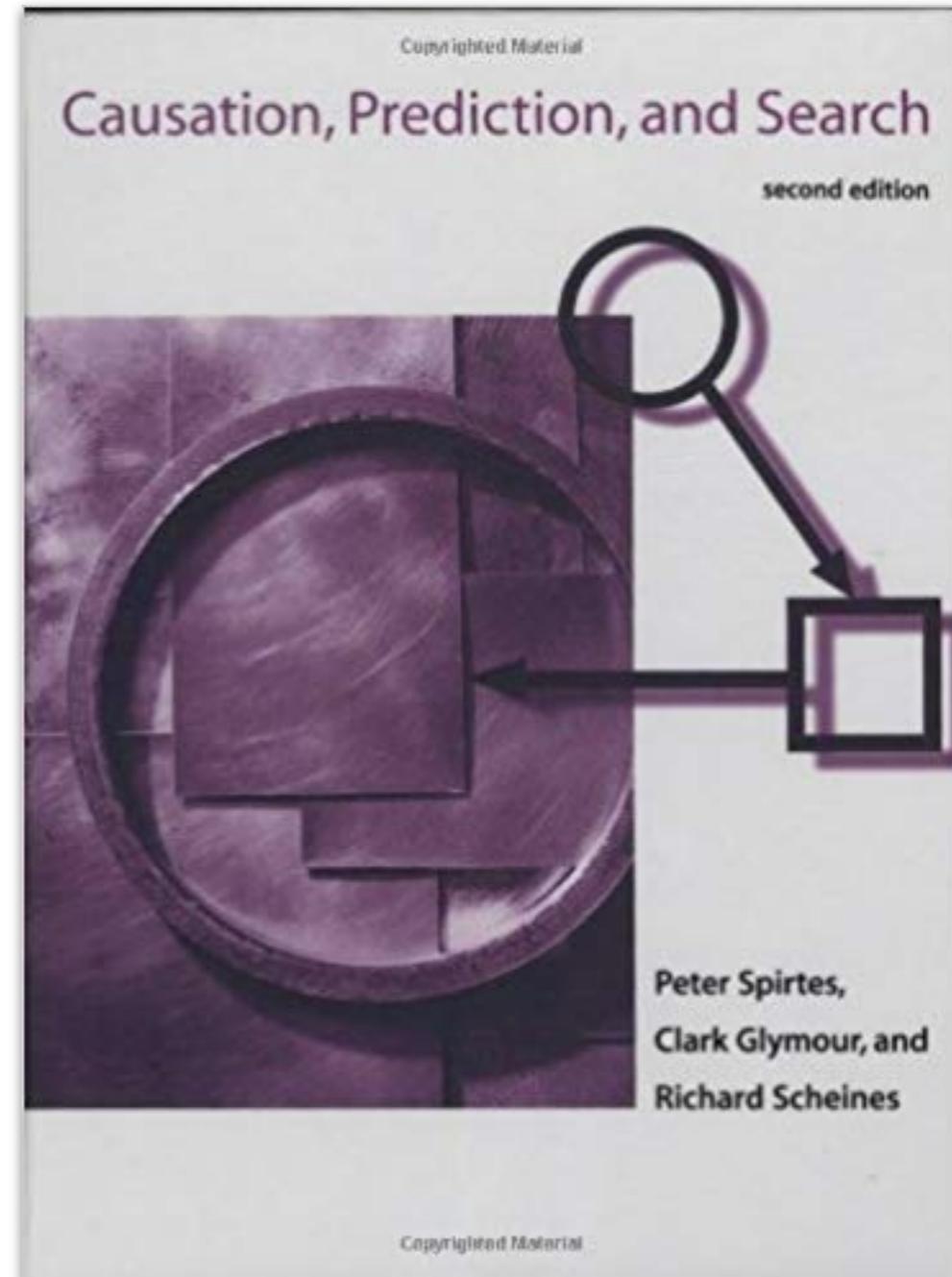
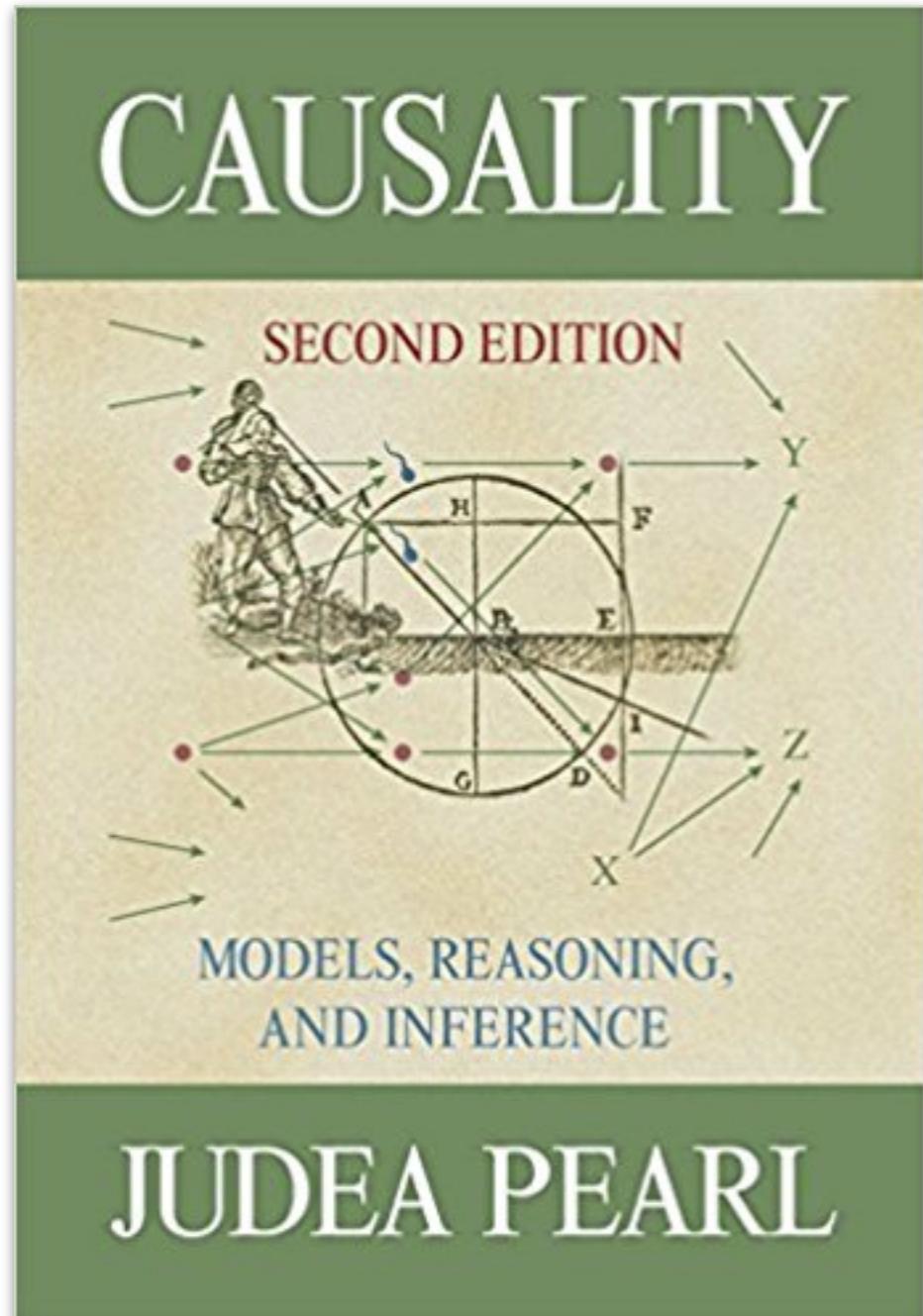
- The Design of Research
- 1 - Designing Research
- 2 - Research Questions
- 3 - Describing Variables
- 4 - Describing Relationships
- 5 - Identification
- 6 - Causal Diagrams
- 7 - Drawing Causal Diagrams
- 8 - Causal Paths and Closing Back Doors
- 9 - Finding Front Doors
- 10 - Treatment Effects
- 11 - Causality with Less Modeling
- The Toolbox
- 12 - Opening the Toolbox
- 13 - Regression
- 14 - Matching
- 15 - Simulation
- 16 - Fixed Effects
- 17 - Event Studies
- 18 - Difference-in-Differences
- 19 - Instrumental Variables
- 20 - Regression Discontinuity
- 21 - A Gallery of Rogues: Other Methods
- 22 - Under the Rug
- References

Click here to [order your copy of The Effect from Chapman & Hall now!](#) It can also be ordered on [Amazon](#) or [Barnes and Noble](#).

Welcome to the web version of *The Effect*. *The Effect* is now out in published form from Chapman & Hall, but they have allowed this free Bookdown version to remain here on theeffectbook.net. This Bookdown version will continue to be free, but I also hope that you will [purchase the published version now that it is available](#). If you would like to be kept informed about updates to the book, such as when new teaching materials come out, please add your email to the [mailing list](#) (no more than one email/month).

made with RMarkdown

<https://theeffectbook.net/>



Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, England: Cambridge University Press.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. The MIT Press. 35

Breakout room

Suppose there is a robust, statistically significant, and long-term correlation between the color of cars and the annual rate at which they are involved in accidents.

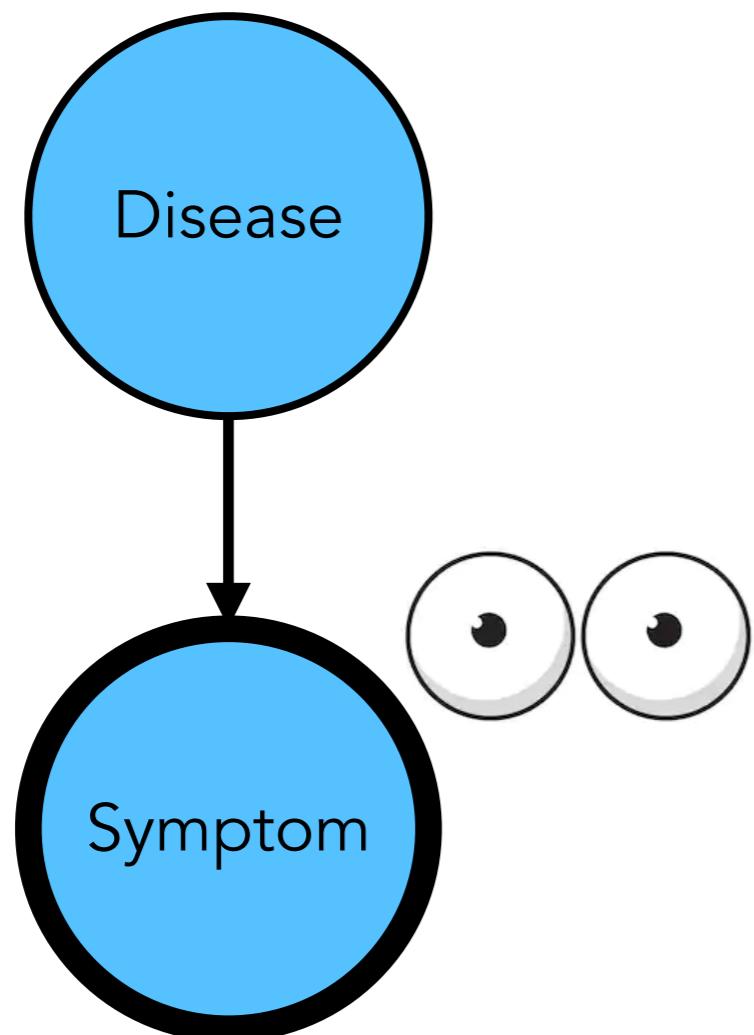
To be concrete, assume that red cars, in particular, are involved in accidents year after year at a higher rate than cars of any other color. When you go to buy a new car, should you avoid the color red in your quest to remain safe on the road?

Hidden

Seeing vs. doing

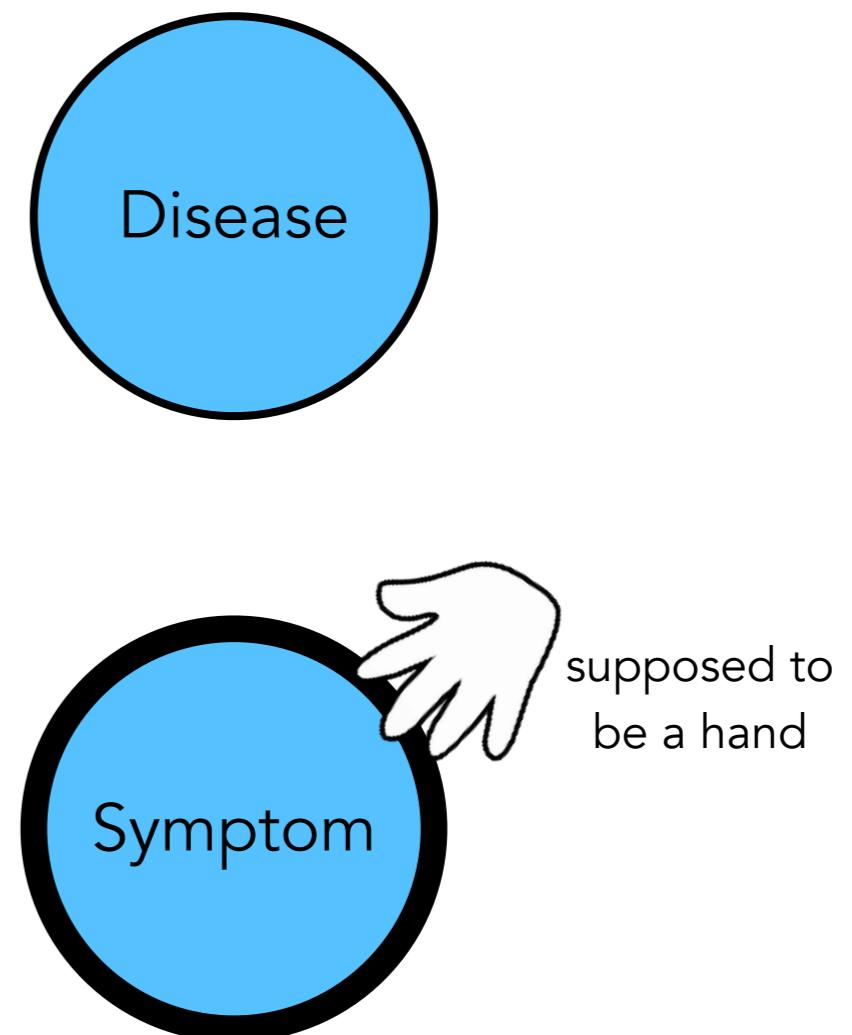
Observation vs. Intervention

seeing



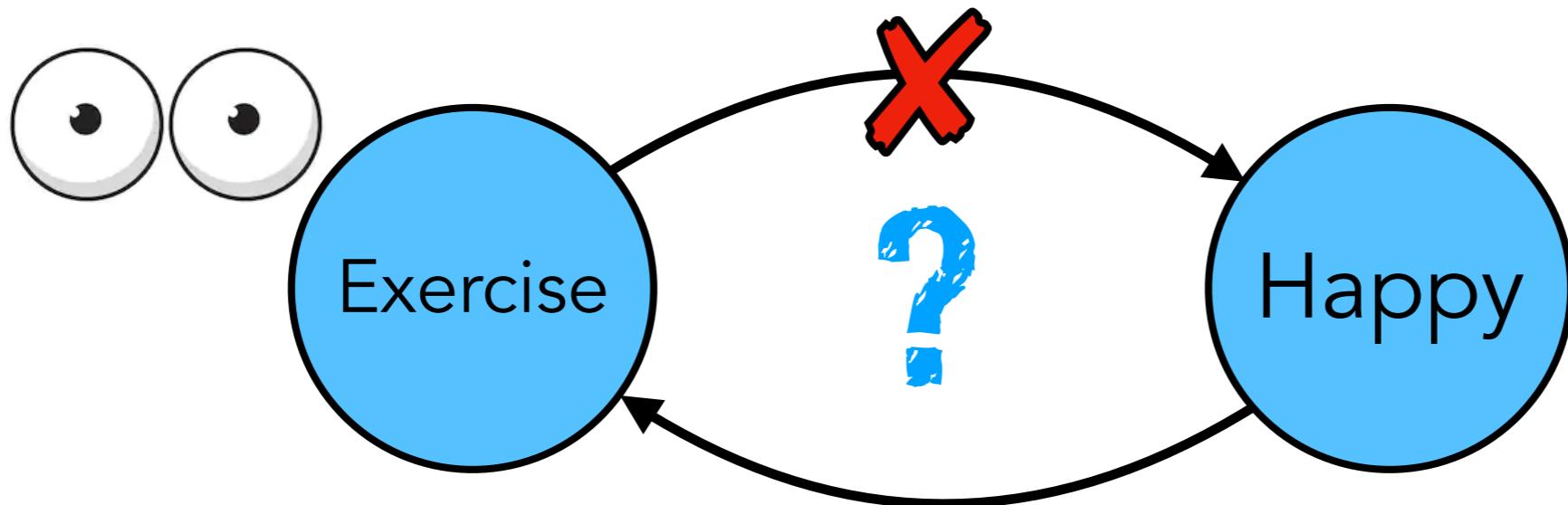
$$p(D | S) > p(D)$$

doing

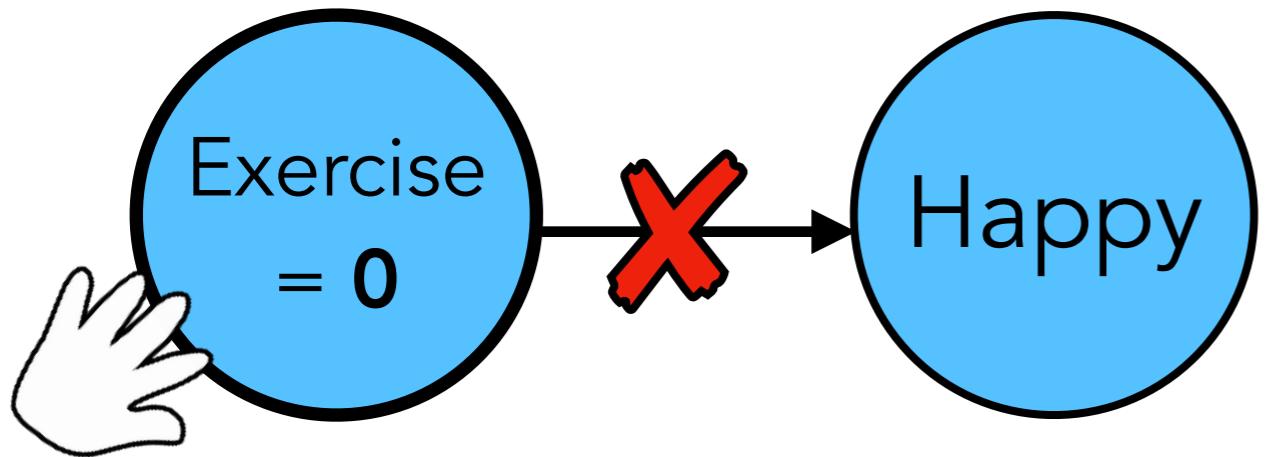


$$p(D | \text{do}(S)) = p(D)$$

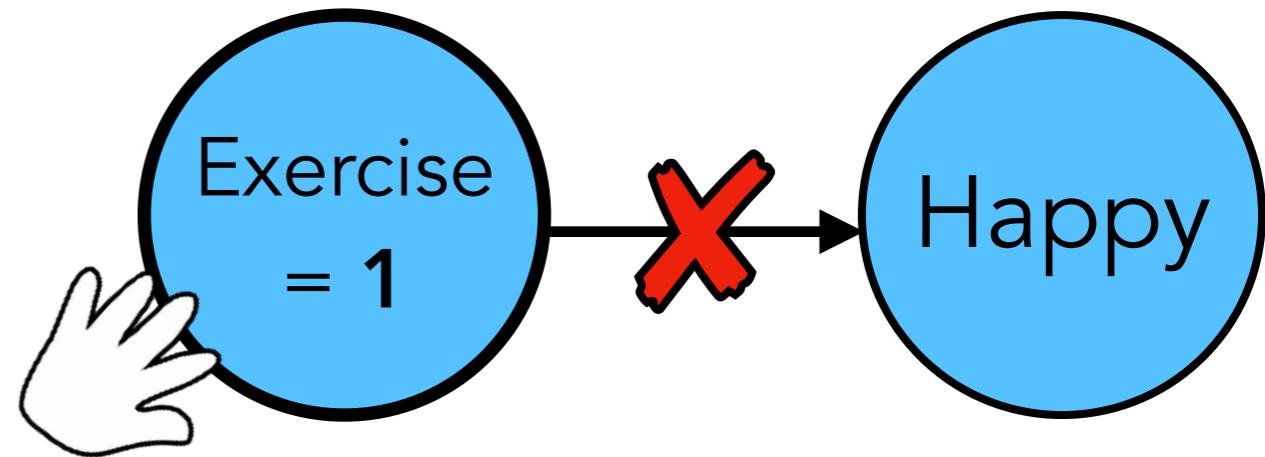
Inferring causal structure through intervention



Experiment 1

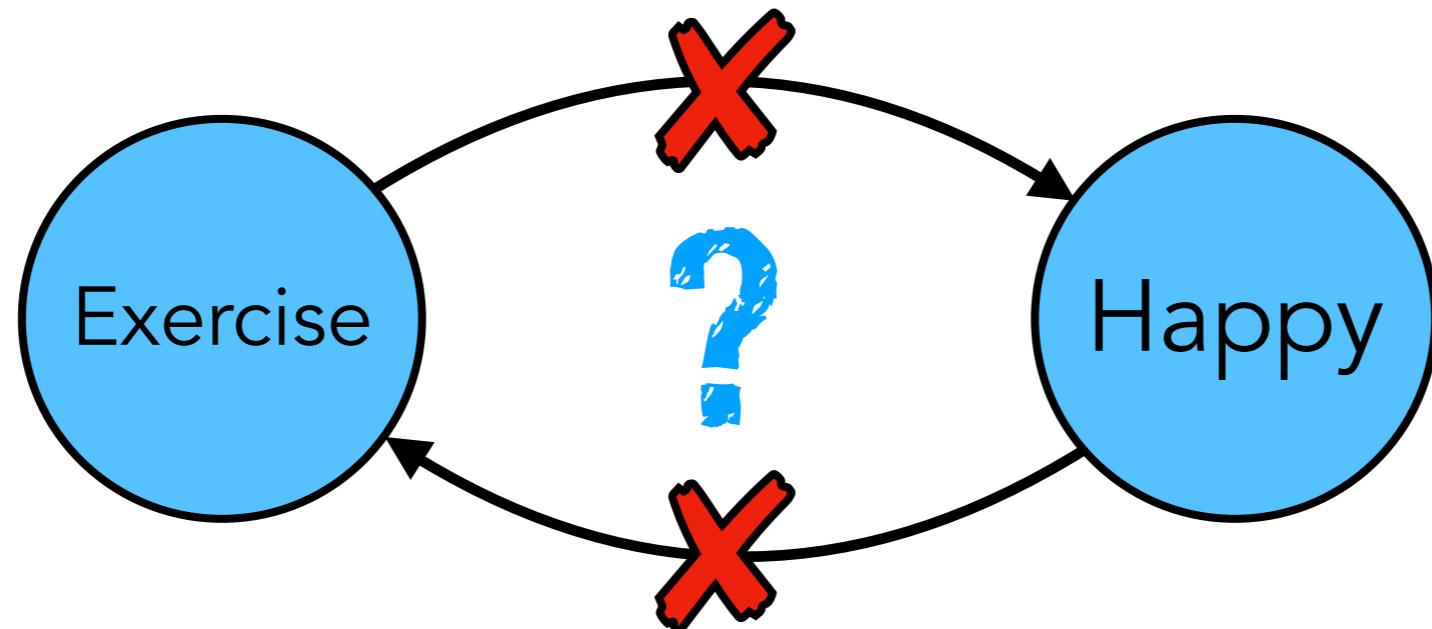


$$p(\text{Happy} \mid \text{do}(\text{Exercise} = 0)) = 0.3$$

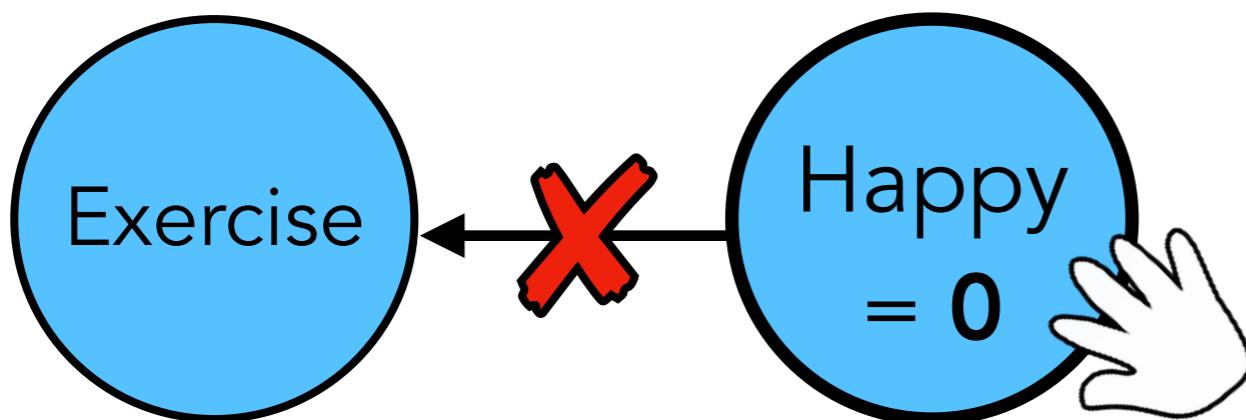


$$p(\text{Happy} \mid \text{do}(\text{Exercise} = 1)) = 0.3$$

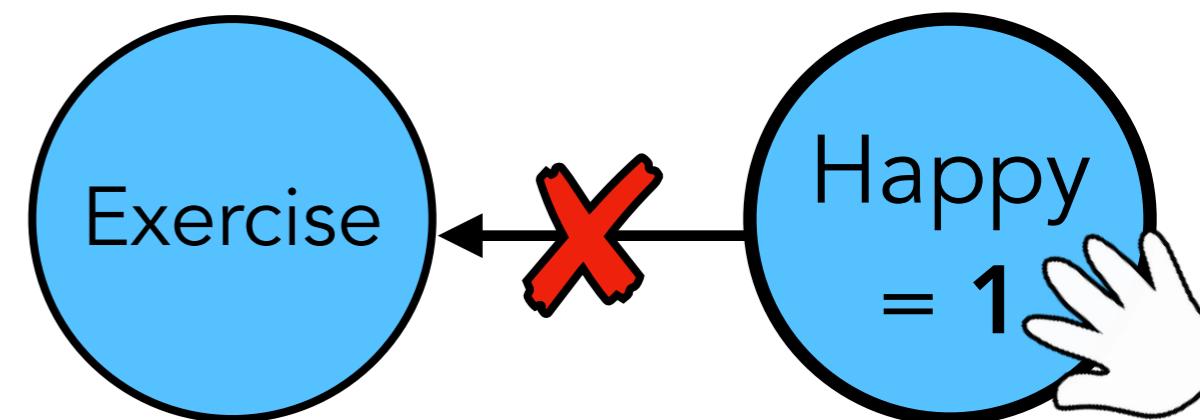
Inferring causal structure through intervention



Experiment 2

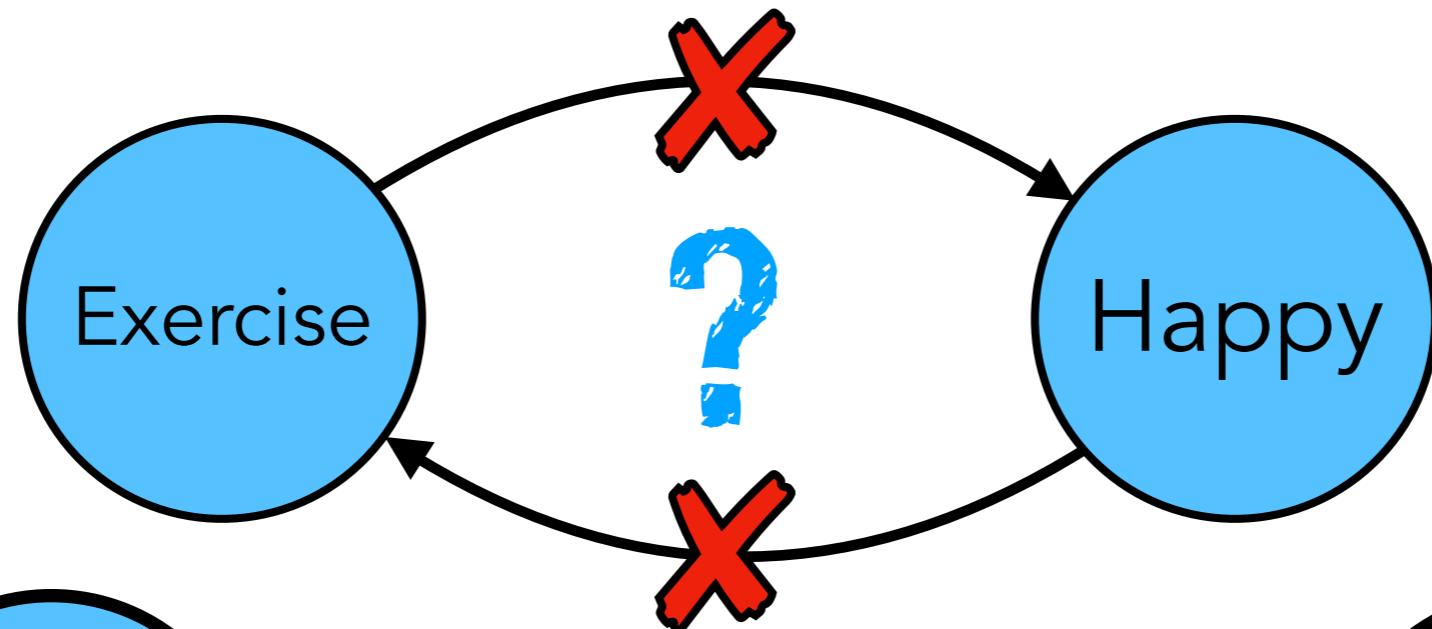


$$p(\text{Exercise} | \text{do}(\text{Happy} = 0)) = 0.1$$

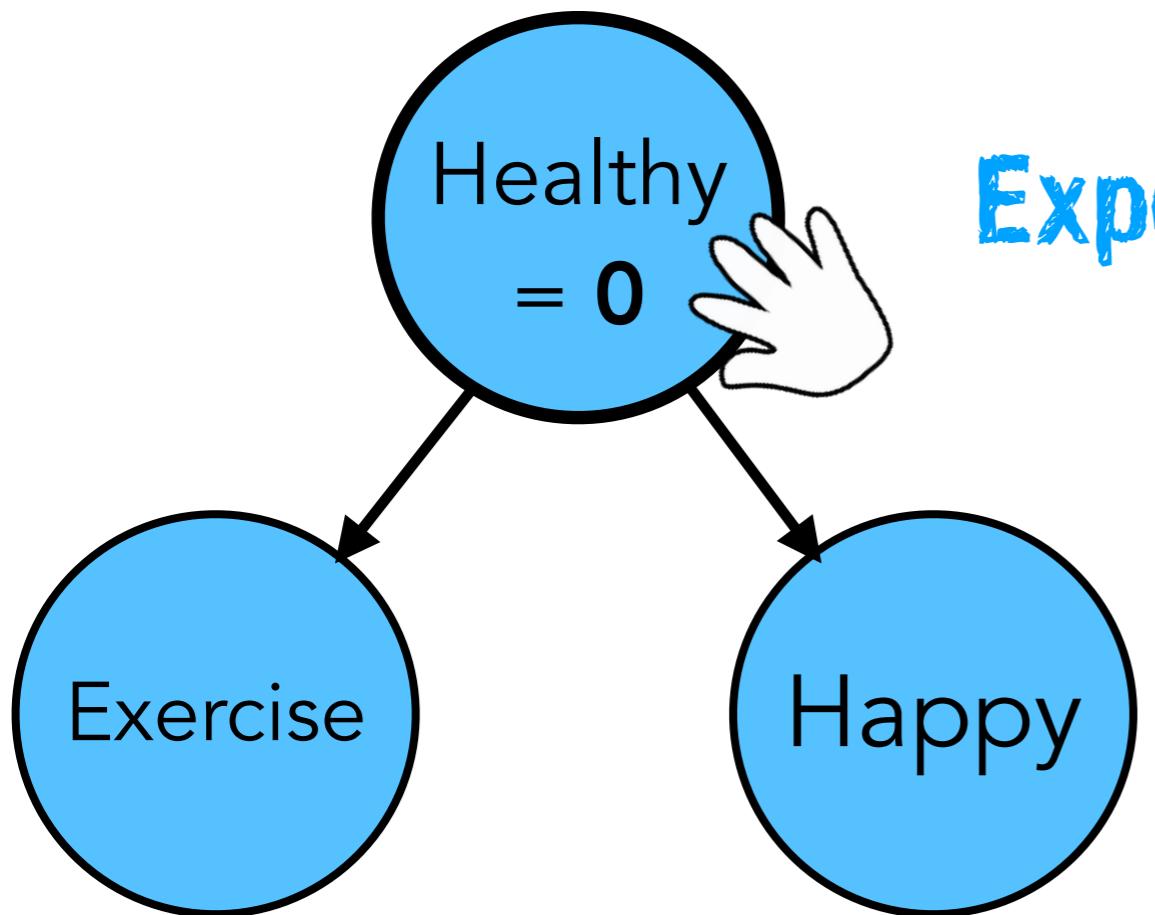


$$p(\text{Exercise} | \text{do}(\text{Happy} = 1)) = 0.1$$

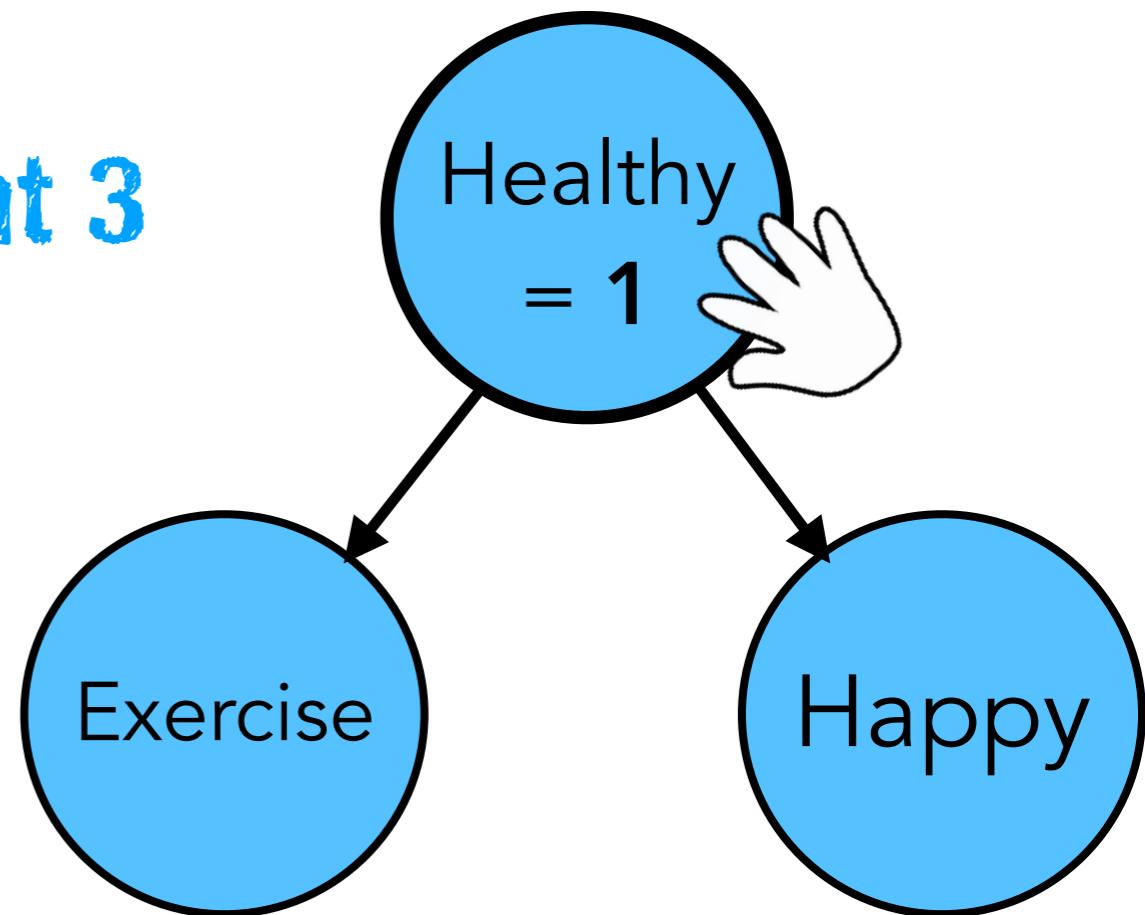
Inferring causal structure through intervention



Experiment 3



$$p(\text{Happy} | \text{do}(\text{Healthy} = 0)) = 0.05$$
$$p(\text{Exercise} | \text{do}(\text{Healthy} = 0)) = 0.1$$

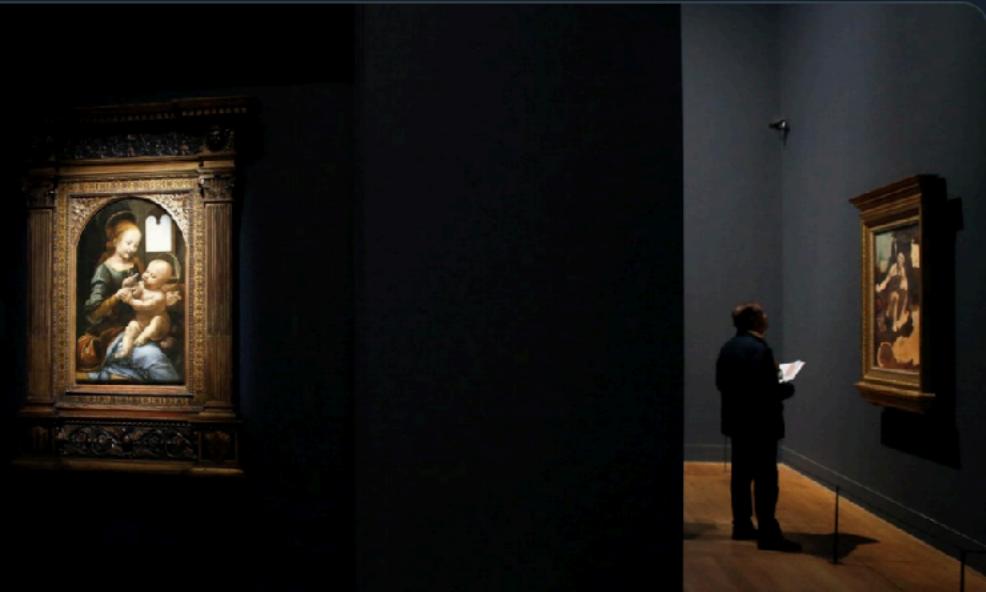


$$p(\text{Happy} | \text{do}(\text{Healthy} = 0)) = 0.5$$
$$p(\text{Exercise} | \text{do}(\text{Healthy} = 0)) = 0.75$$



NYT Health
@NYTHealth

Want to live longer? Try going to the opera. Researchers in Britain have found that people who reported going to a museum or concert even once a year lived longer than those who didn't.



Another Benefit to Going to Museums? You May Live Longer

Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.

[nytimes.com](#)

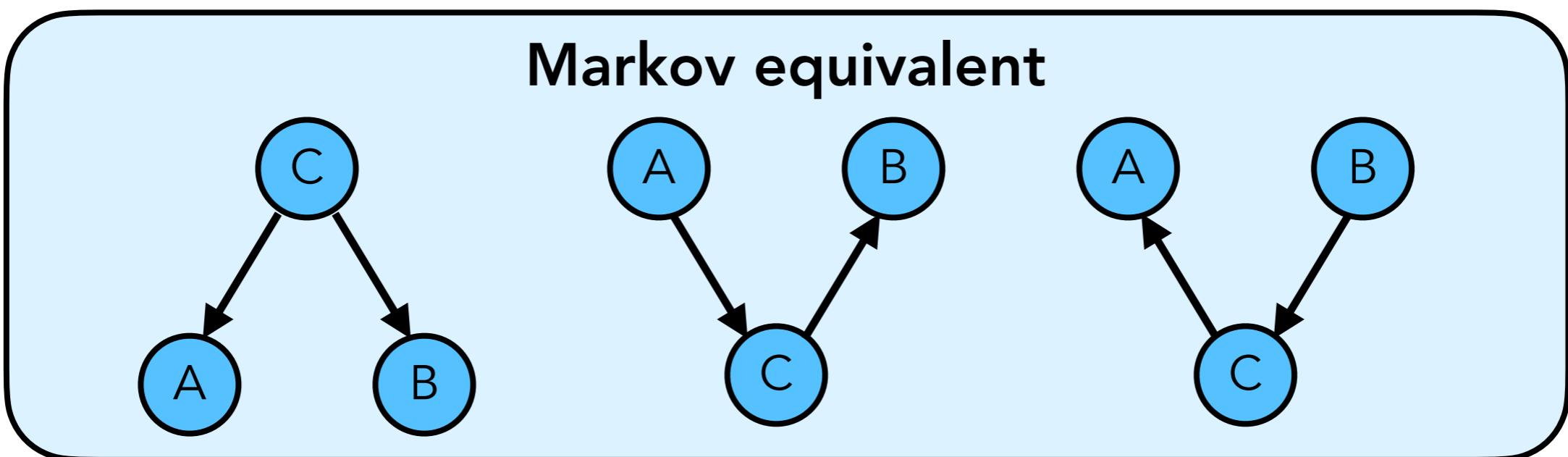
9:19 AM · Dec 22, 2019 · SocialFlow

336 Retweets 1.3K Likes



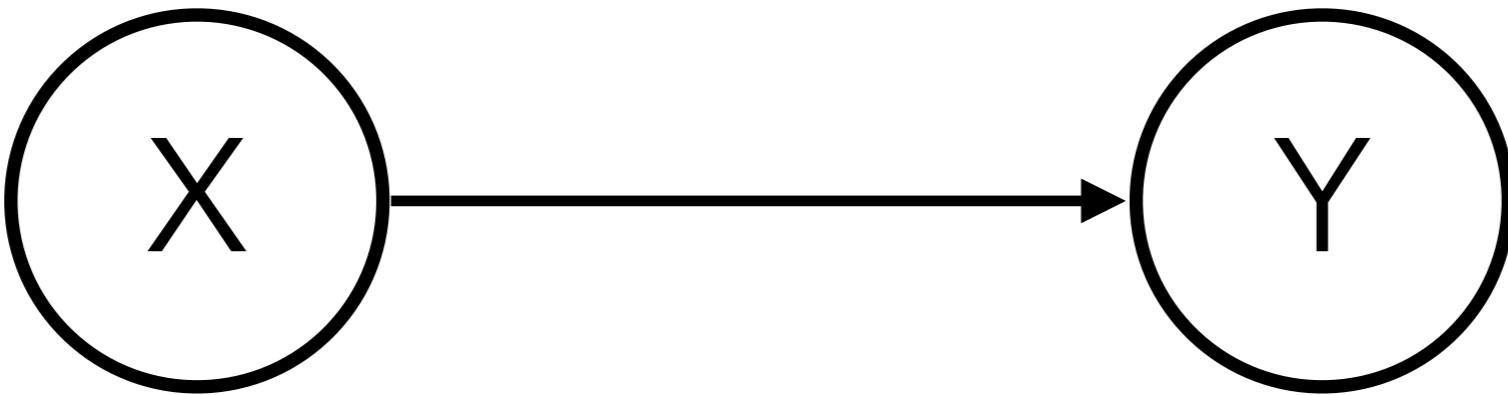
Important take home message

- correlation is not causation
- correlation (= probabilistic dependence)
suggests that there is some causal relationship
- but we don't know which one it is

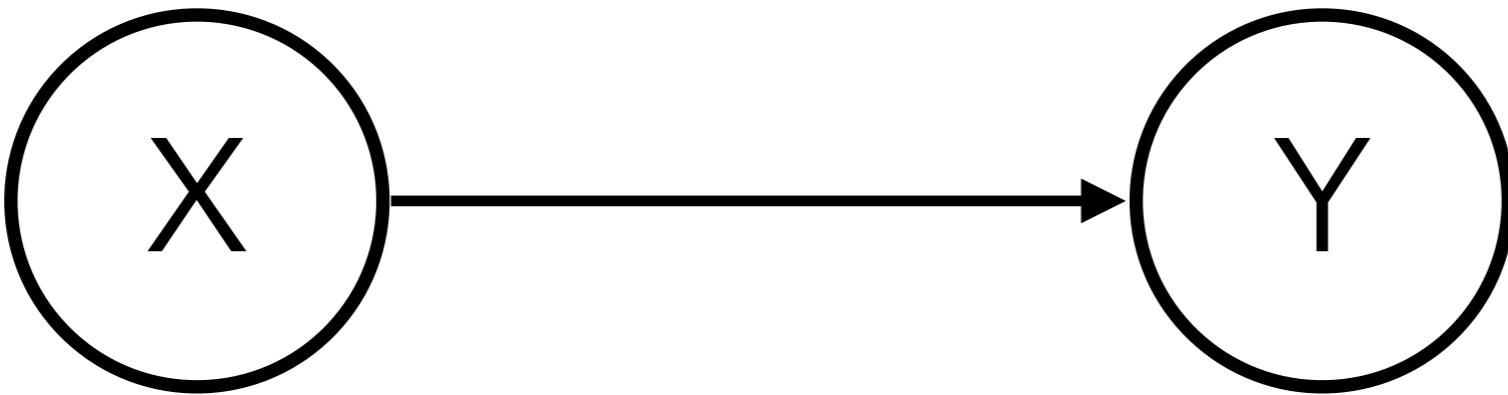


- **causal interventions** / experiments can reveal the underlying causal structure

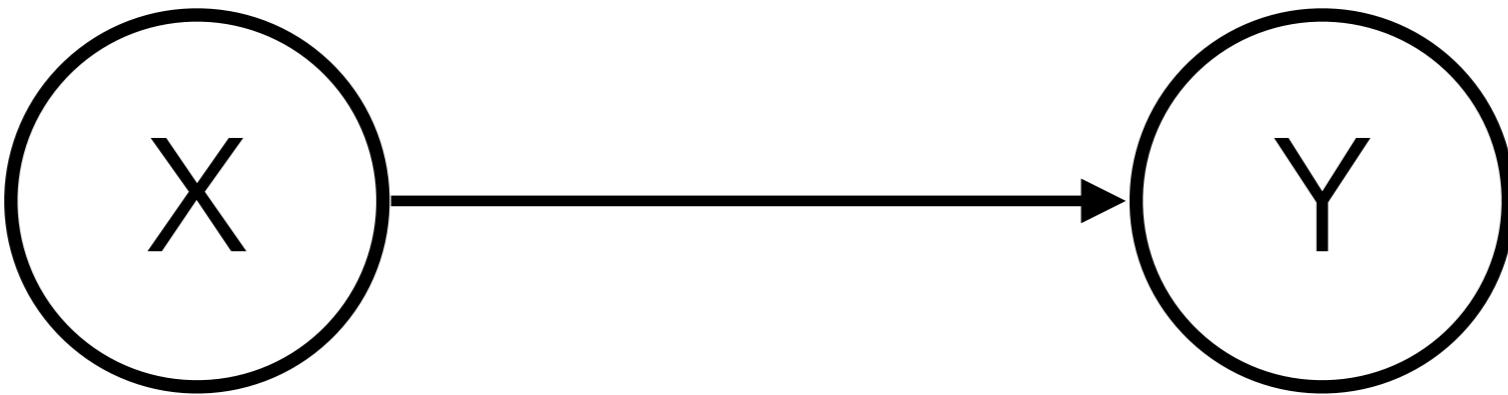
Observation, intervention, counterfactual



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 <i>found</i> the sprinkler off?



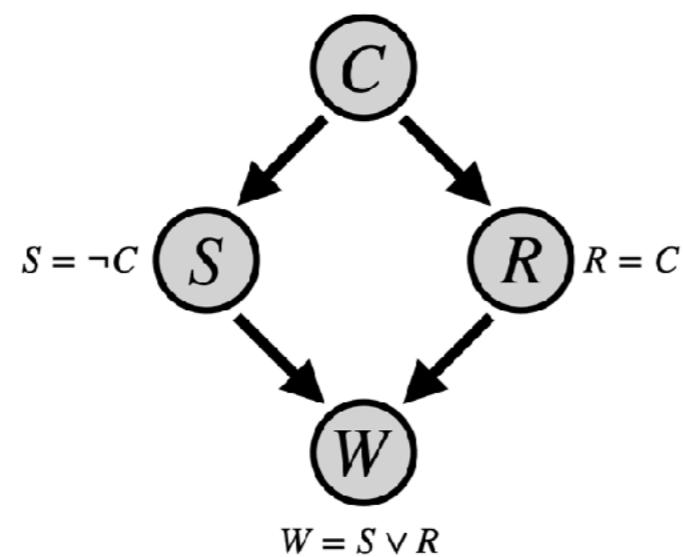
Level	Concept	Expression	Activity	Question	Example
I	Observation/ Prediction	$p(y x)$	Seeing	How does x change my belief in y ?	Would the grass be wet if we <i>found</i> the sprinkler off?
II	Intervention/ Hypothetical	$p(y \text{do}(x))$	Doing	Would y happen if I did x ?	Would the grass be wet if <i>made sure</i> that the sprinkler was off?



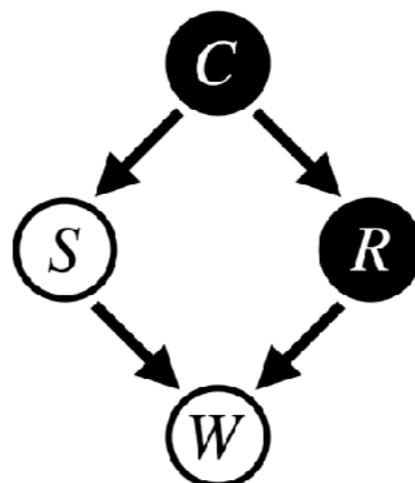
Level	Concept	Expression	Activity	Question	Example
I	Observation/ Prediction	$p(y x)$	Seeing	How does x change my belief in y ?	Would the grass be wet if we <i>found</i> the sprinkler off?
II	Intervention/ Hypothetical	$p(y \text{do}(x))$	Doing	Would y happen if I did x ?	Would the grass be wet if <i>made sure</i> 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 <i>had made sure</i> that the sprinkler was off, given that the grass is wet and the sprinkler on?

A Causal structure

$$p(C = \text{true}) = 0.5$$



B What actually happened



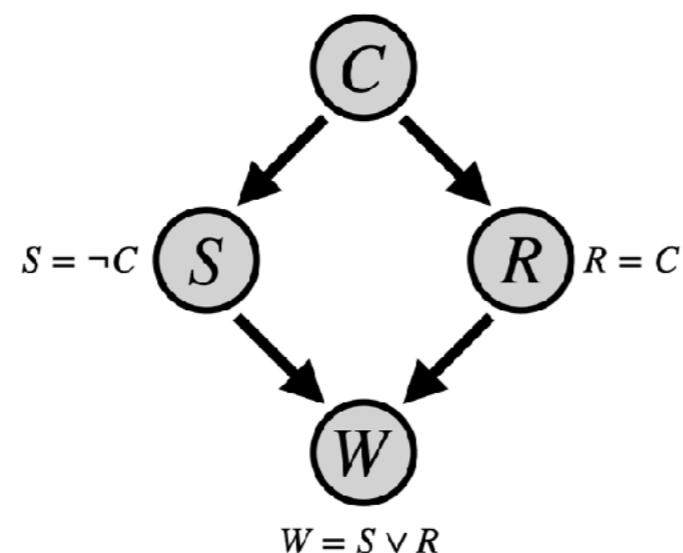
There were no **clouds**,
it didn't **rain**, the
sprinkler was **on**, and
the grass was **wet**.

Did the sprinkler cause
the grass to be wet?

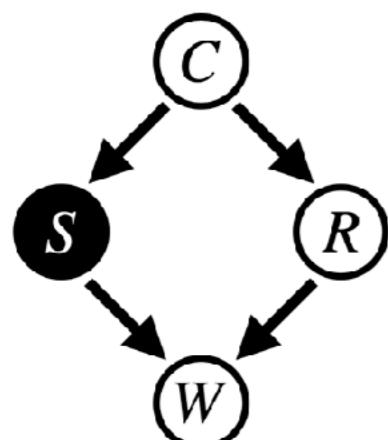
$$p(s \rightarrow w) = ?$$

A Causal structure

$$p(C = \text{true}) = 0.5$$



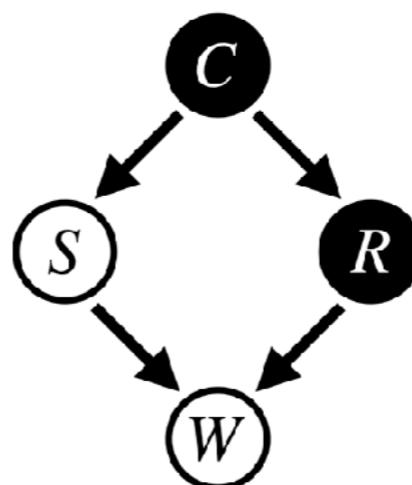
C Observation (Level I)



Is the grass dry when the sprinkler is off?

$$p(w' | s') = 0$$

B What actually happened



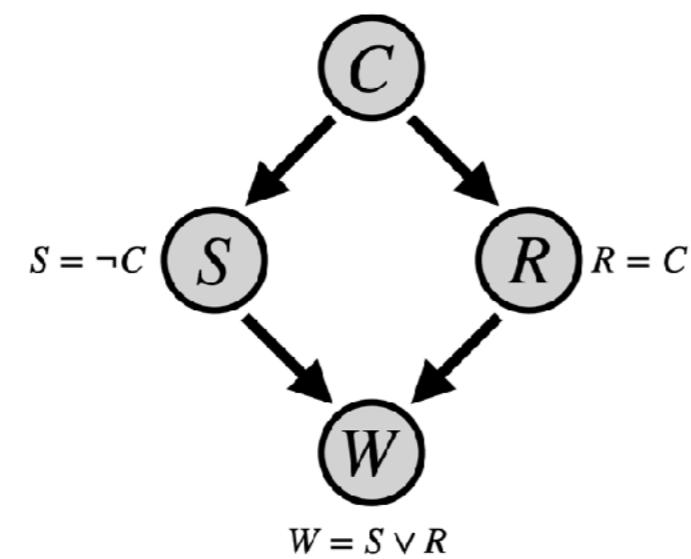
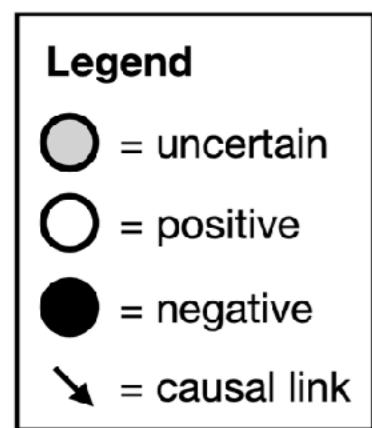
There were no **clouds**, it didn't **rain**, the **sprinkler** was **on**, and the grass was **wet**.

Did the sprinkler cause the grass to be wet?

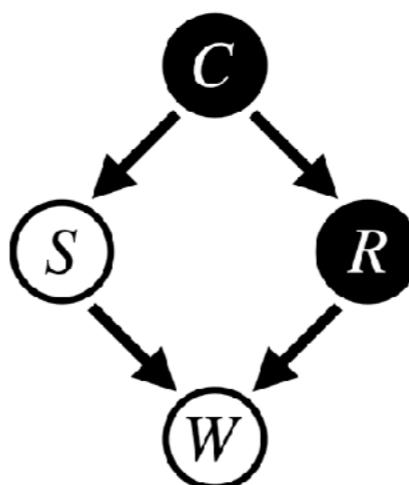
$$p(s \rightarrow w) = ?$$

A Causal structure

$$p(C = \text{true}) = 0.5$$



B What actually happened

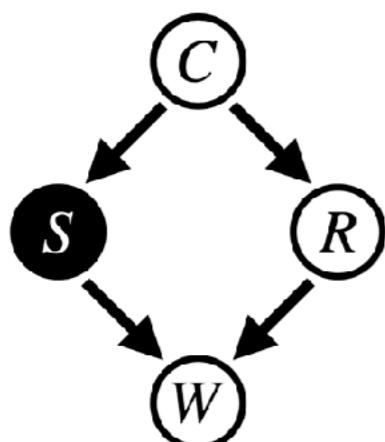


There were no **clouds**, it didn't **rain**, the **sprinkler** was on, and the grass was **wet**.

Did the sprinkler cause the grass to be wet?

$$p(s \rightarrow w) = ?$$

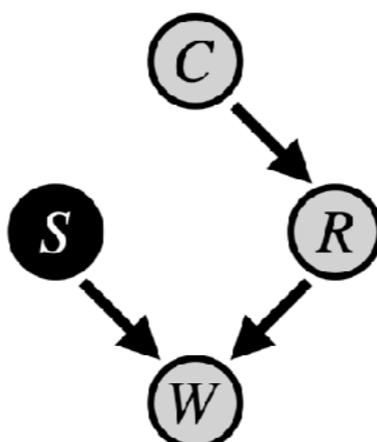
C Observation (Level I)



Is the grass dry when the sprinkler is off?

$$p(w' | s') = 0$$

D Intervention (Level II)

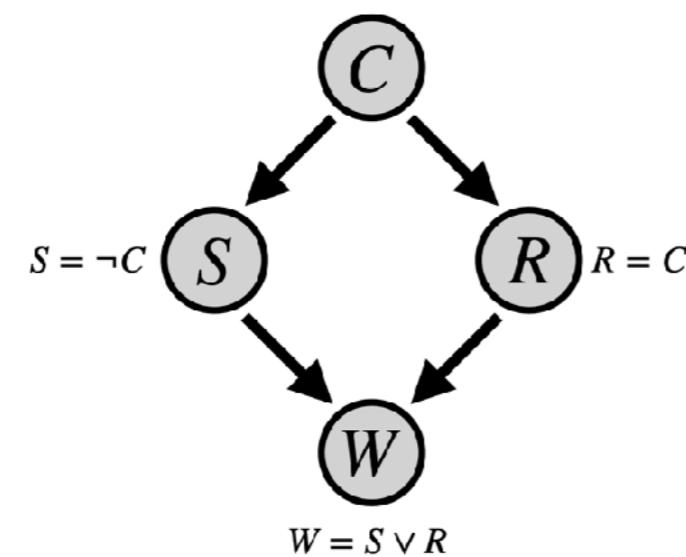
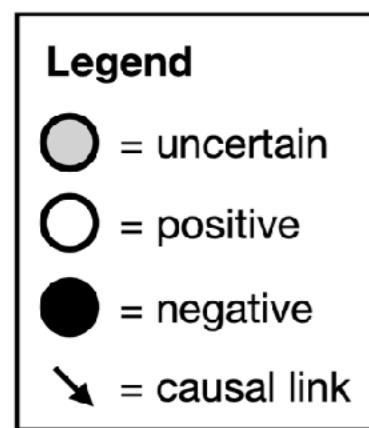


Would the grass be dry if the sprinkler was off?

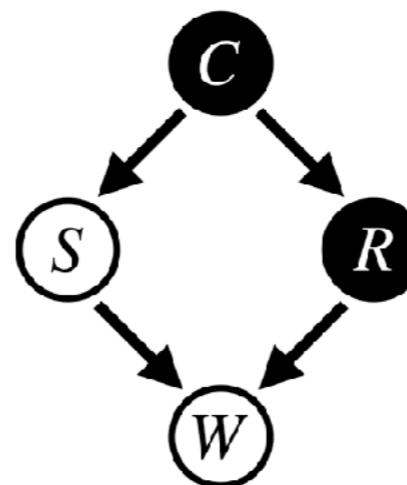
$$p(w' | \text{do}(s')) = 0.5$$

A Causal structure

$$p(C = \text{true}) = 0.5$$



B What actually happened

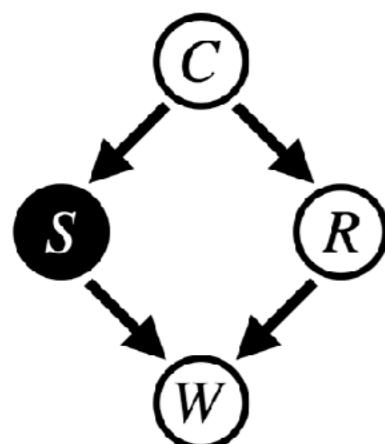


There were no clouds, it didn't rain, the sprinkler was on, and the grass was wet.

Did the sprinkler cause the grass to be wet?

$$p(s \rightarrow w) = ?$$

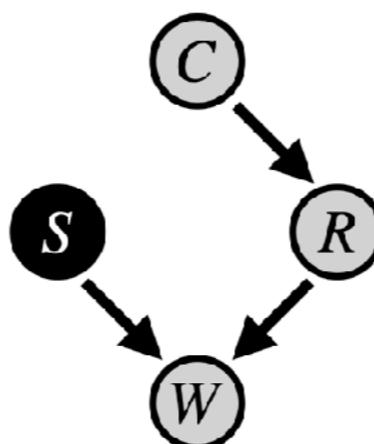
C Observation (Level I)



Is the grass dry when the sprinkler is off?

$$p(w' | s') = 0$$

D Intervention (Level II)

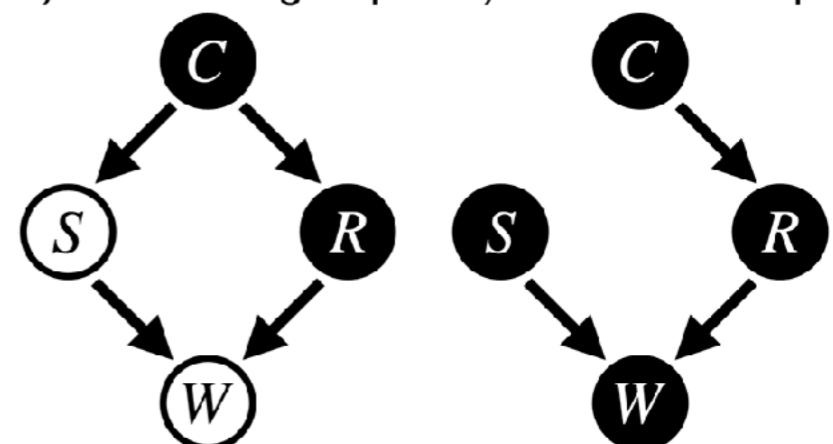


Would the grass be dry if the sprinkler was off?

$$p(w' | \text{do}(s')) = 0.5$$

E Counterfactual (Level III)

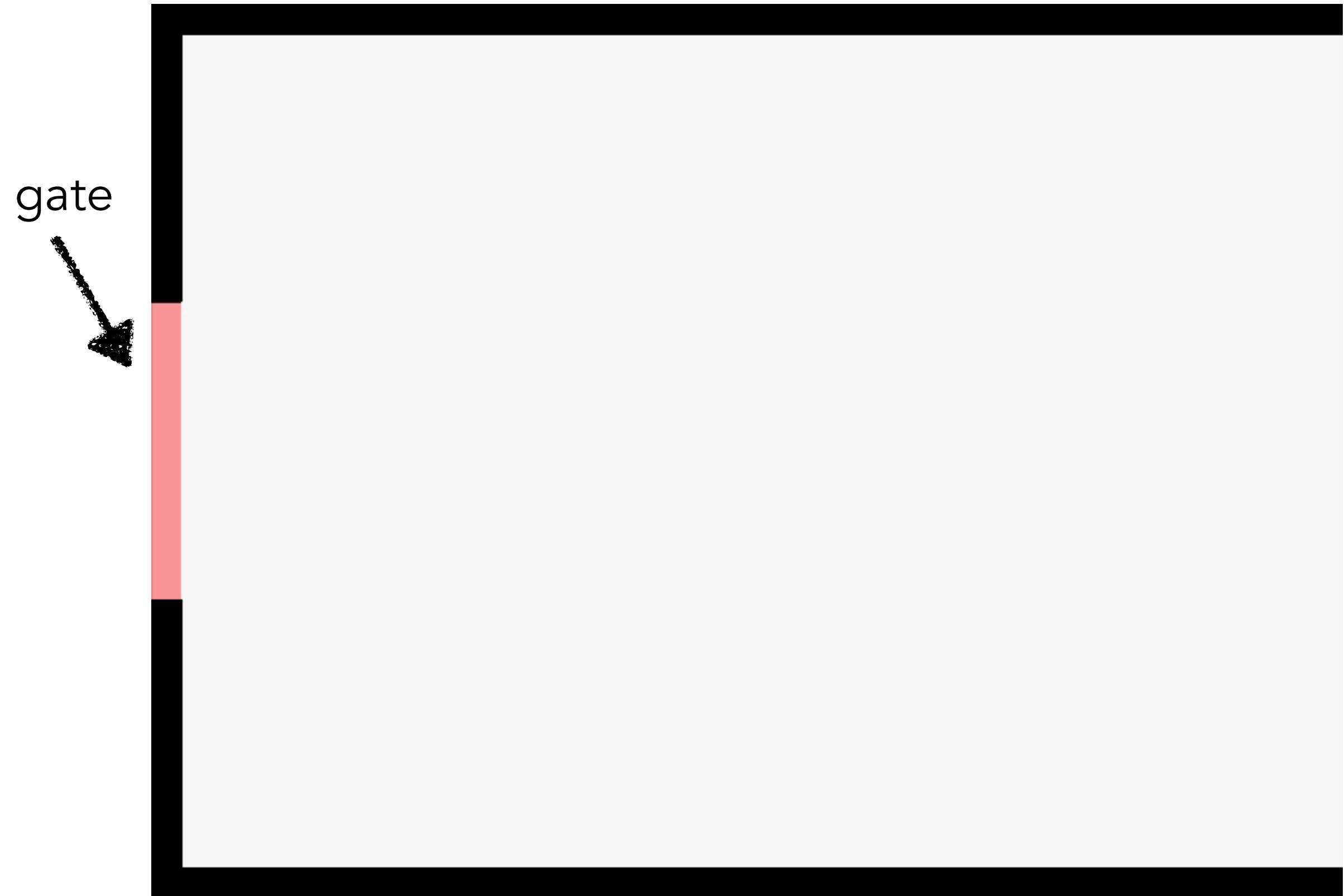
- 1) Conditioning step
- 2) Intervention step



Would the grass have been dry if the sprinkler had been off?

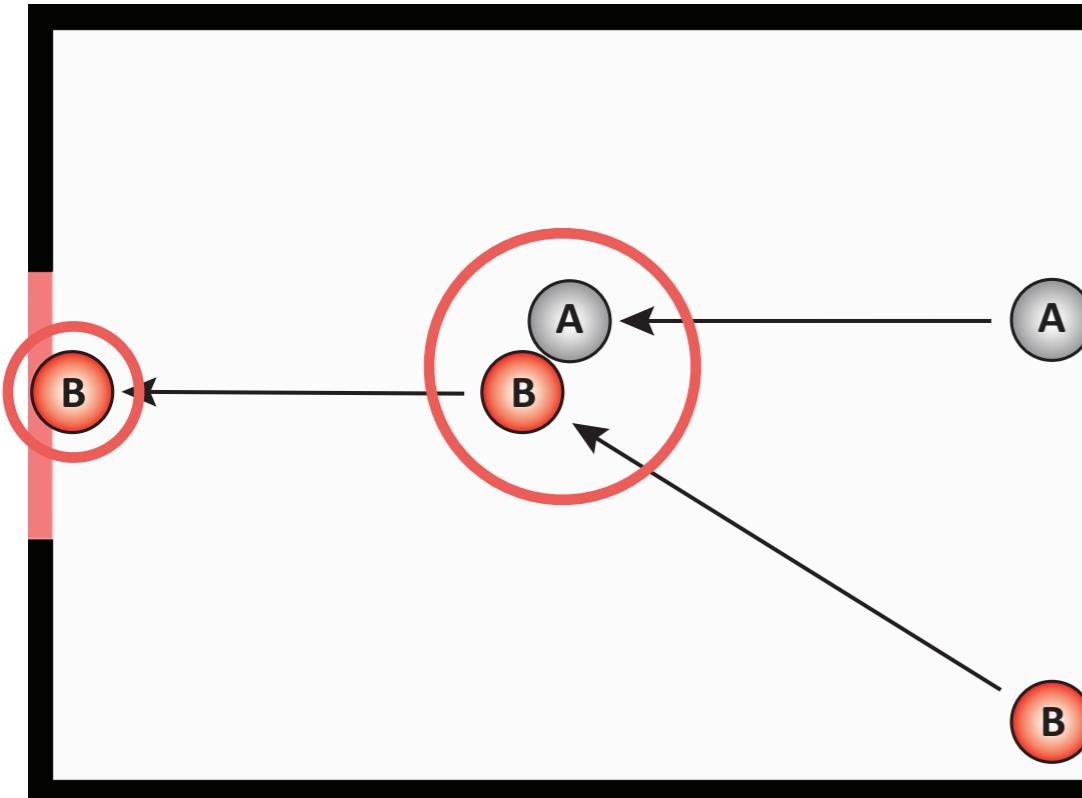
$$p(w'_{s'} | c', s, r', w) = 1$$

Did A cause B to go through the gate?



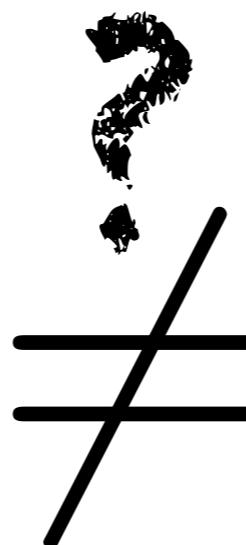
Counterfactual Simulation Model

What happened?

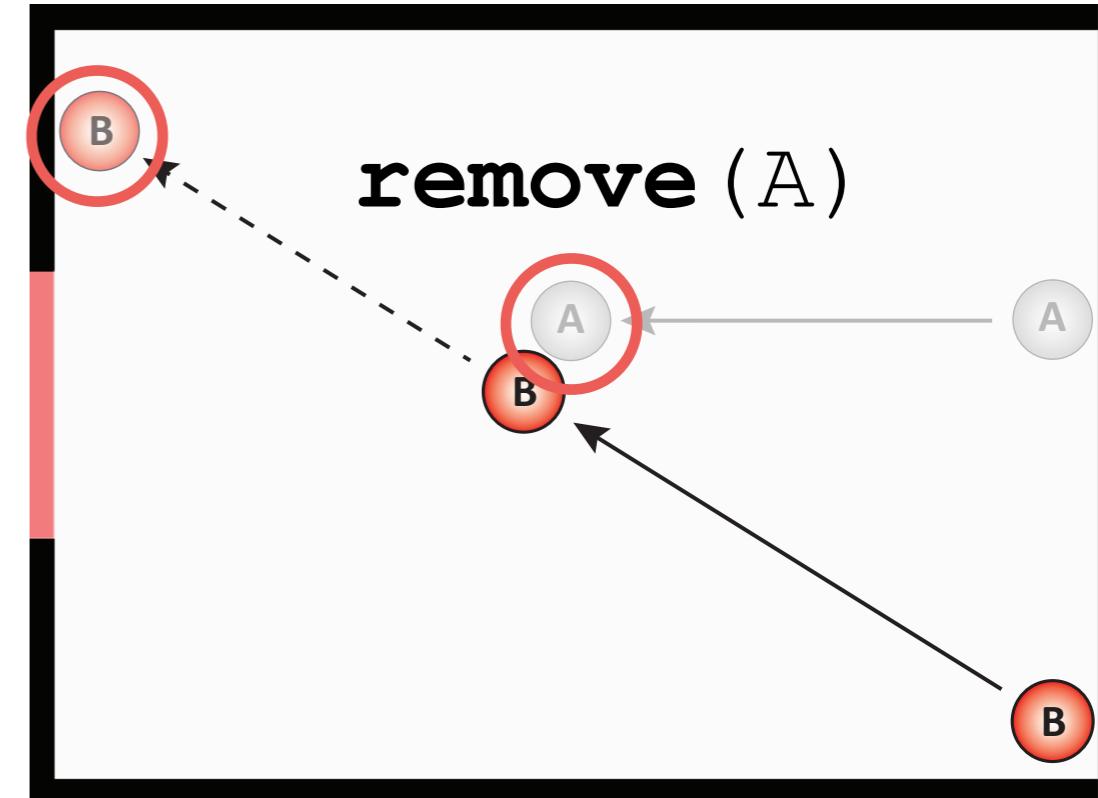


Actual situation

 went through the gate



What would have happened?

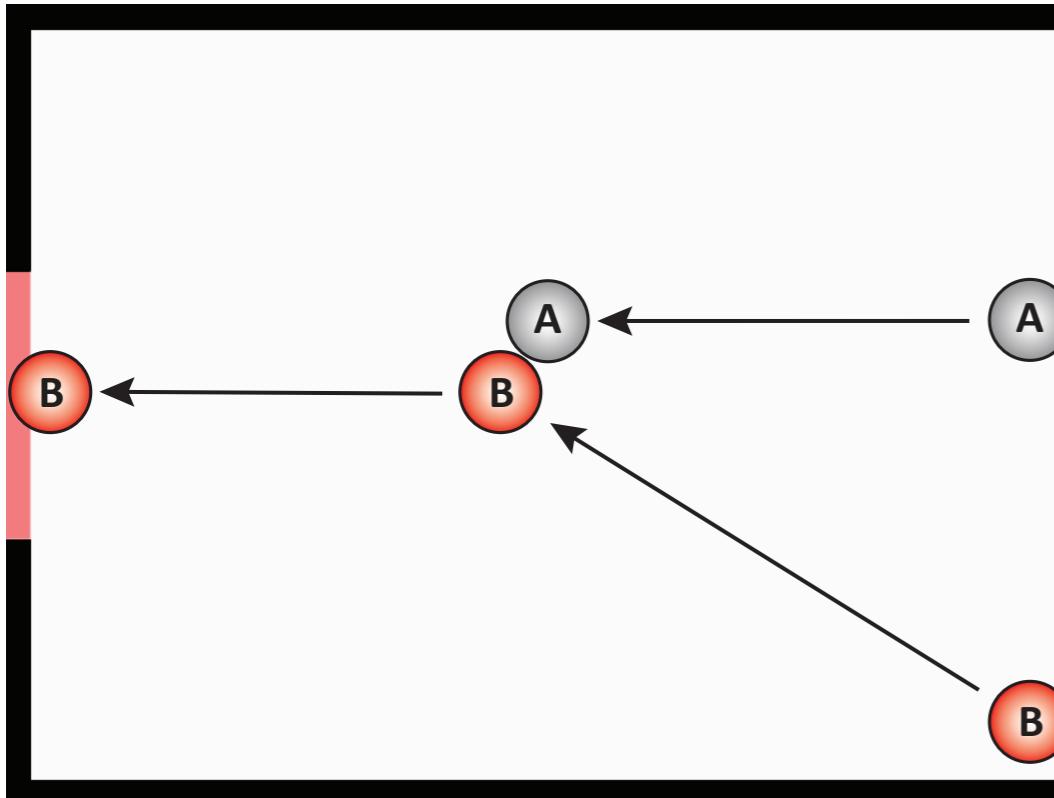


Counterfactual situation

 would have missed the gate

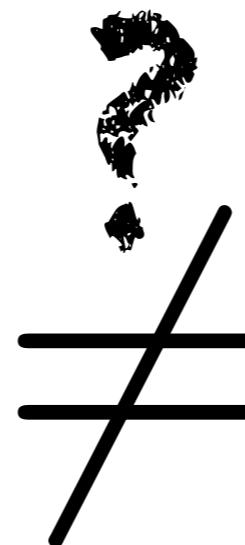
Quantitative predictions

What happened?

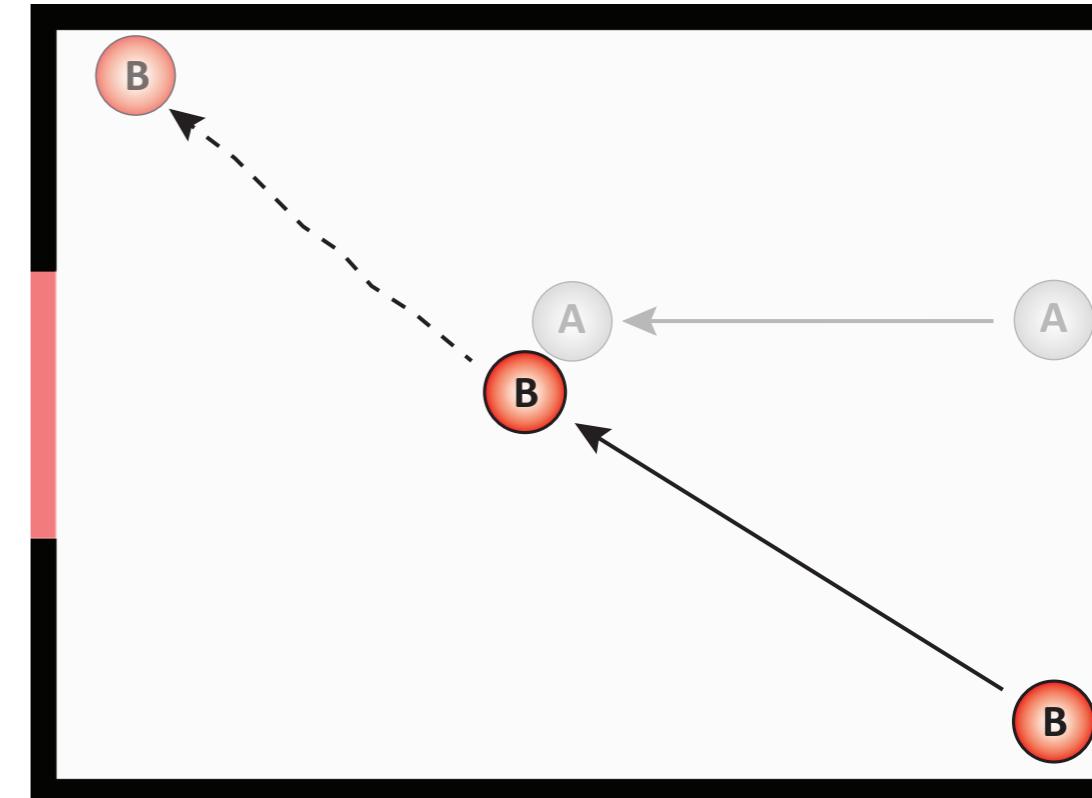


Actual situation

- (B) went through the gate



What would have happened?

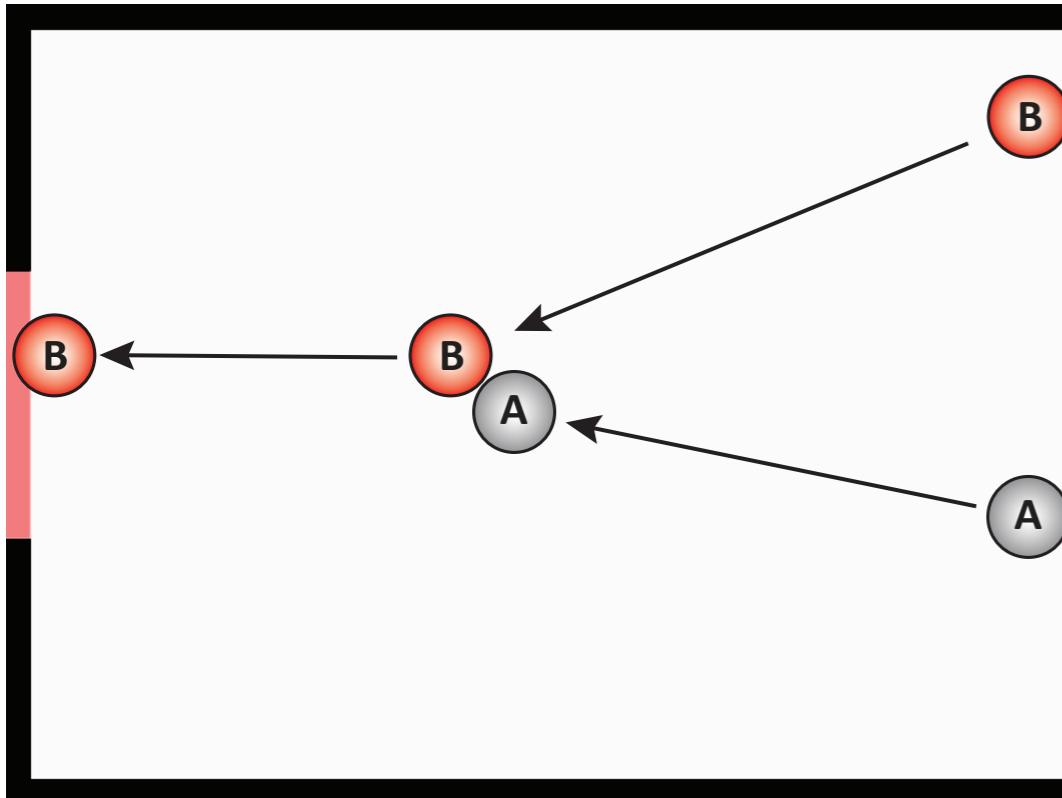


Counterfactual situation

- (B) would have missed the gate ✓
- (B) would have missed the gate ✓
- (B) would have missed the gate ✓

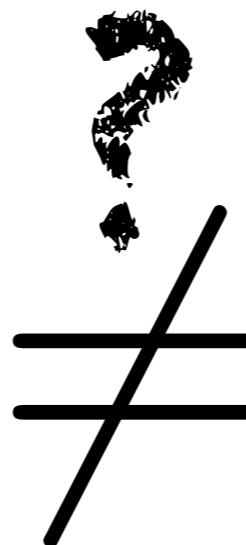
Quantitative predictions

What happened?

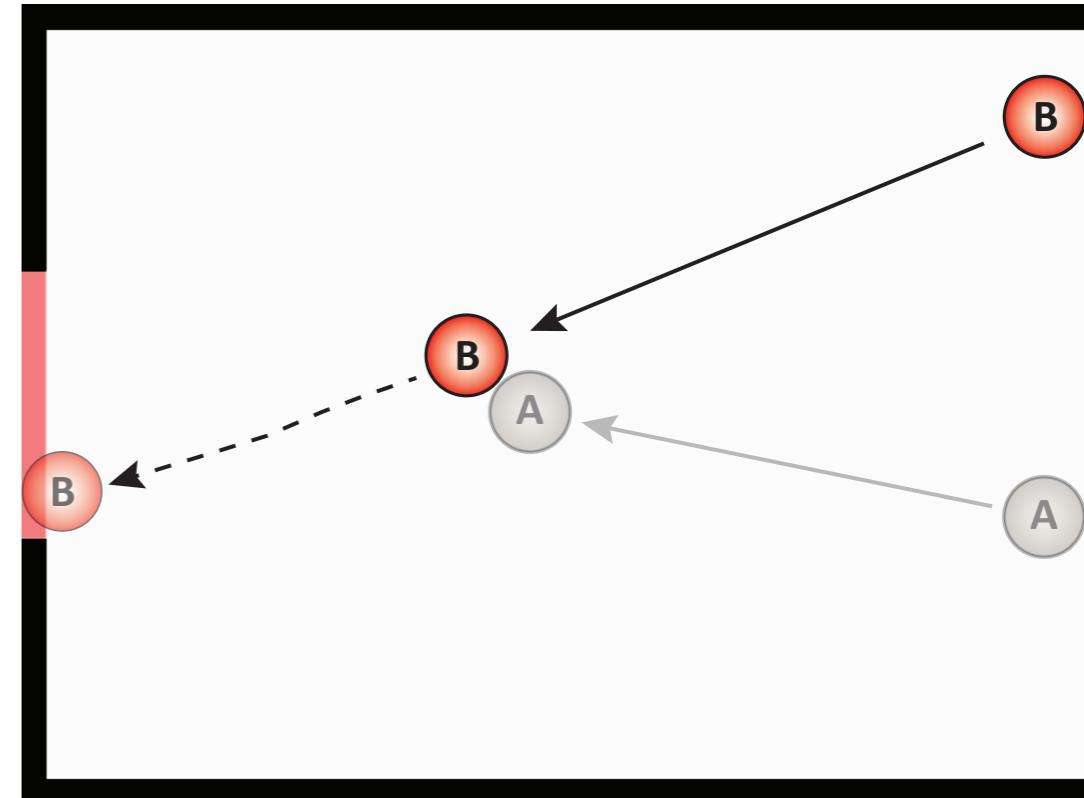


Actual situation

- (B) went through the gate



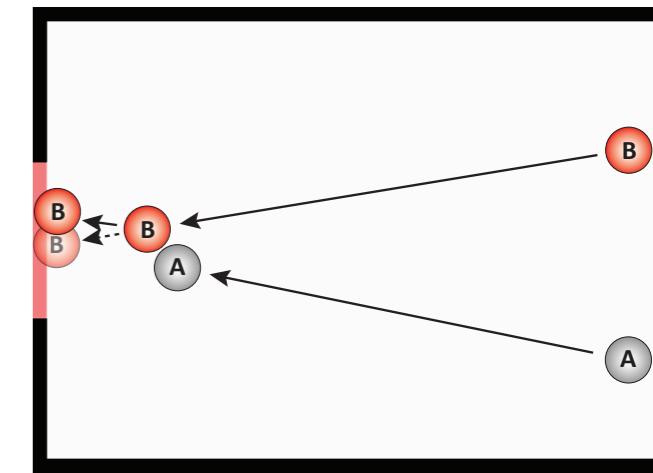
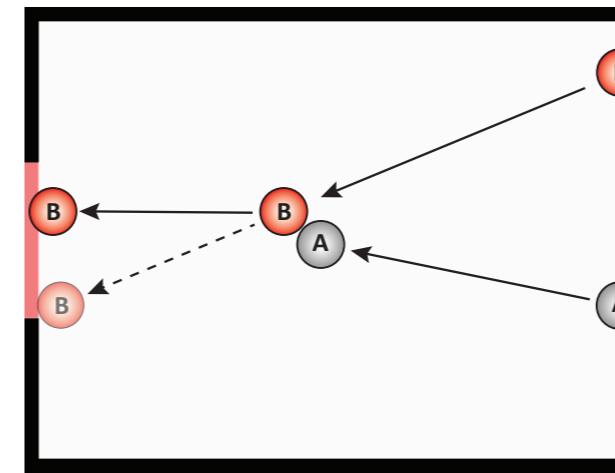
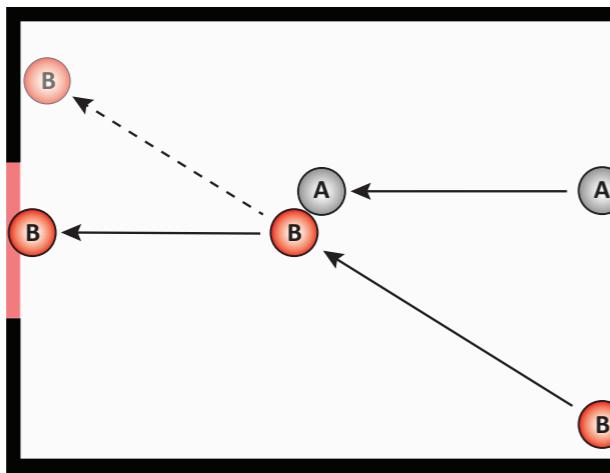
What would have happened?



Counterfactual situation

- (B) would have **missed** the gate ✓
- (B) would have **gone through** gate ✗
- (B) would have **gone through** gate ✗

Quantitative predictions

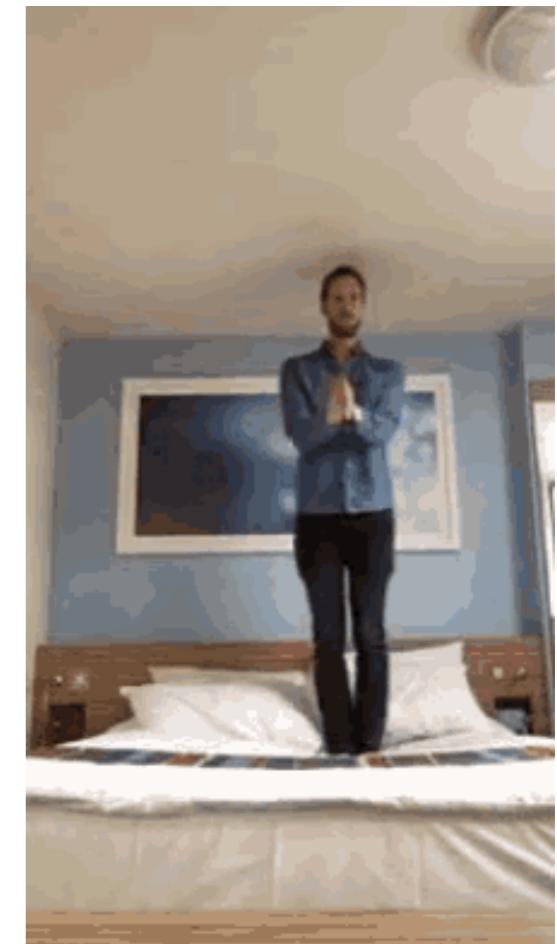
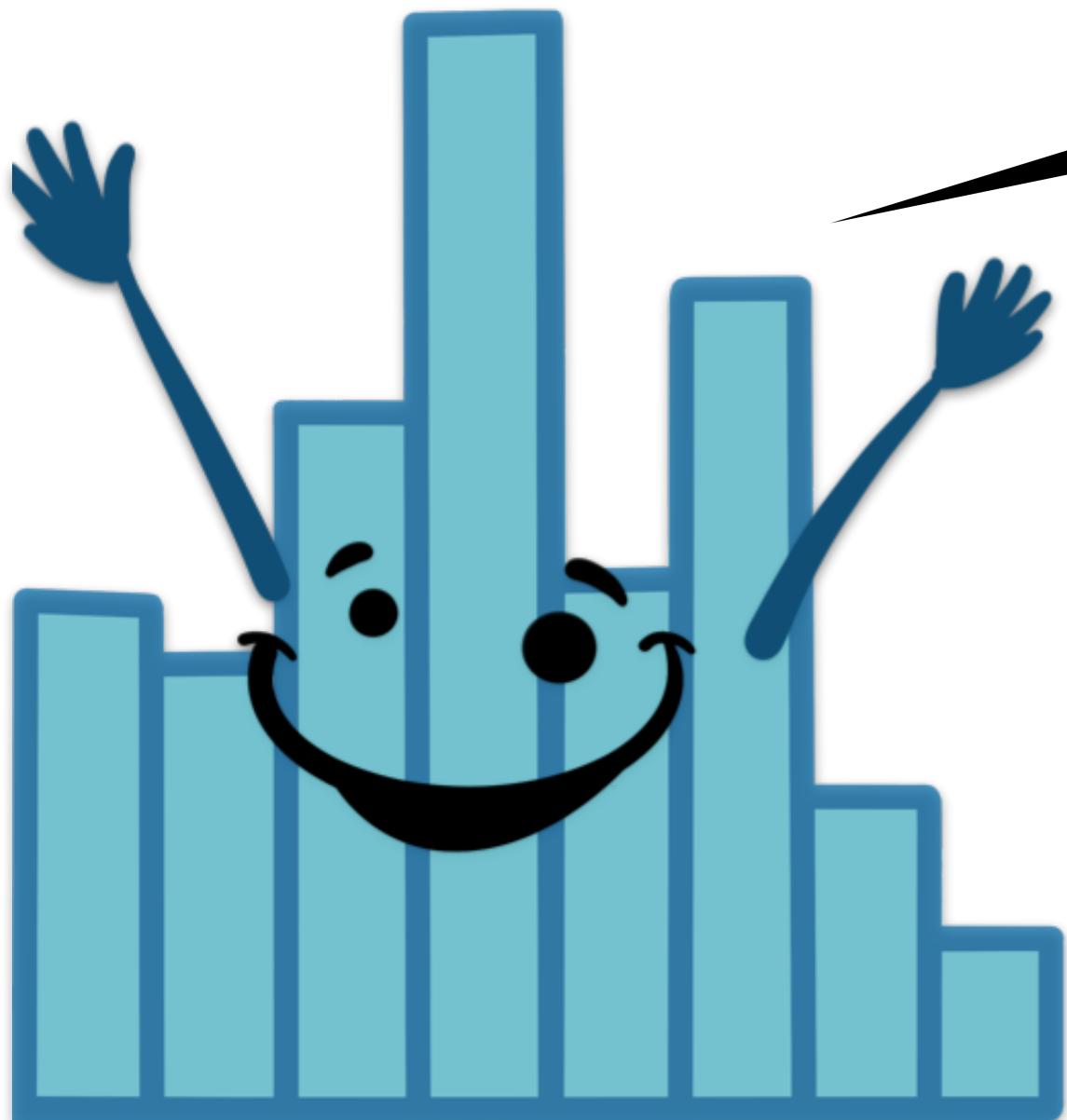


Did A prevent B from go through the gate?

1/2 speed

02:00

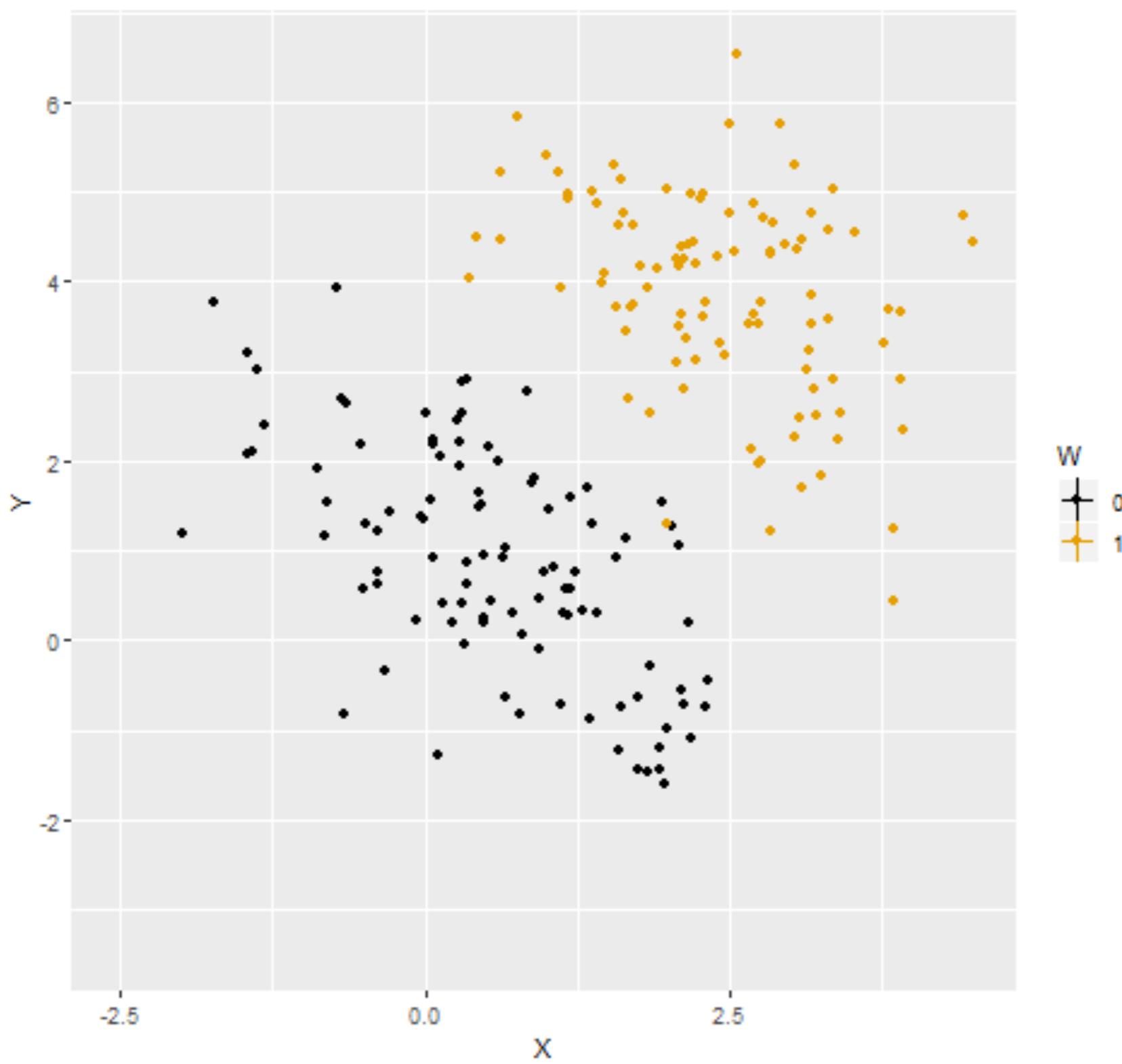
stretch break!



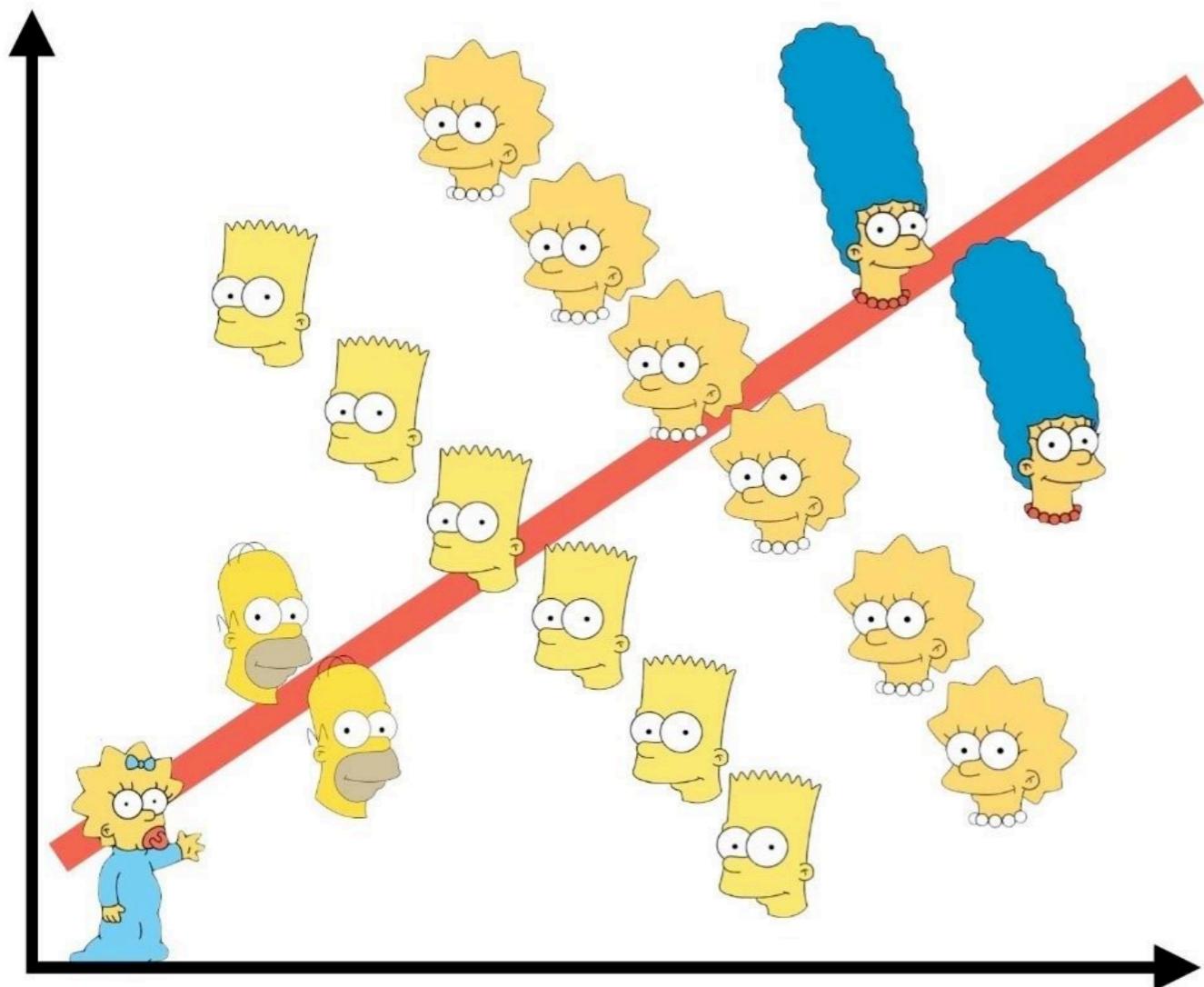
Controlling for variables

The Relationship between Y and X, Controlling for a Binary Variable W

1. Start with raw data. Correlation between X and Y: 0.319



Simpson's paradox

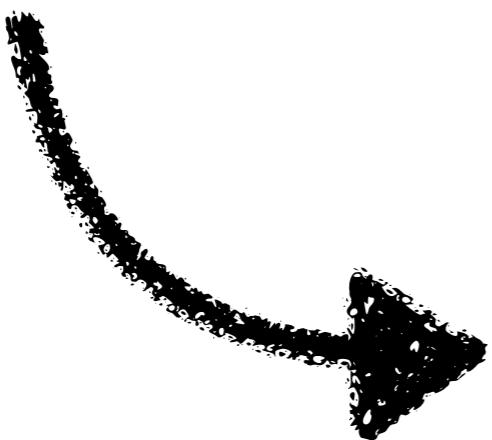


- when the relationship between two variables changes strongly after conditioning on another variable
- interesting real world cases
- **google it!**

What does controlling for variables mean?

we are not actually "**controlling**" the variable

instead, we are taking the variable into consideration when making predictions



the hope is that we get a better estimate of the parameter that we are interested in by taking into account other factors

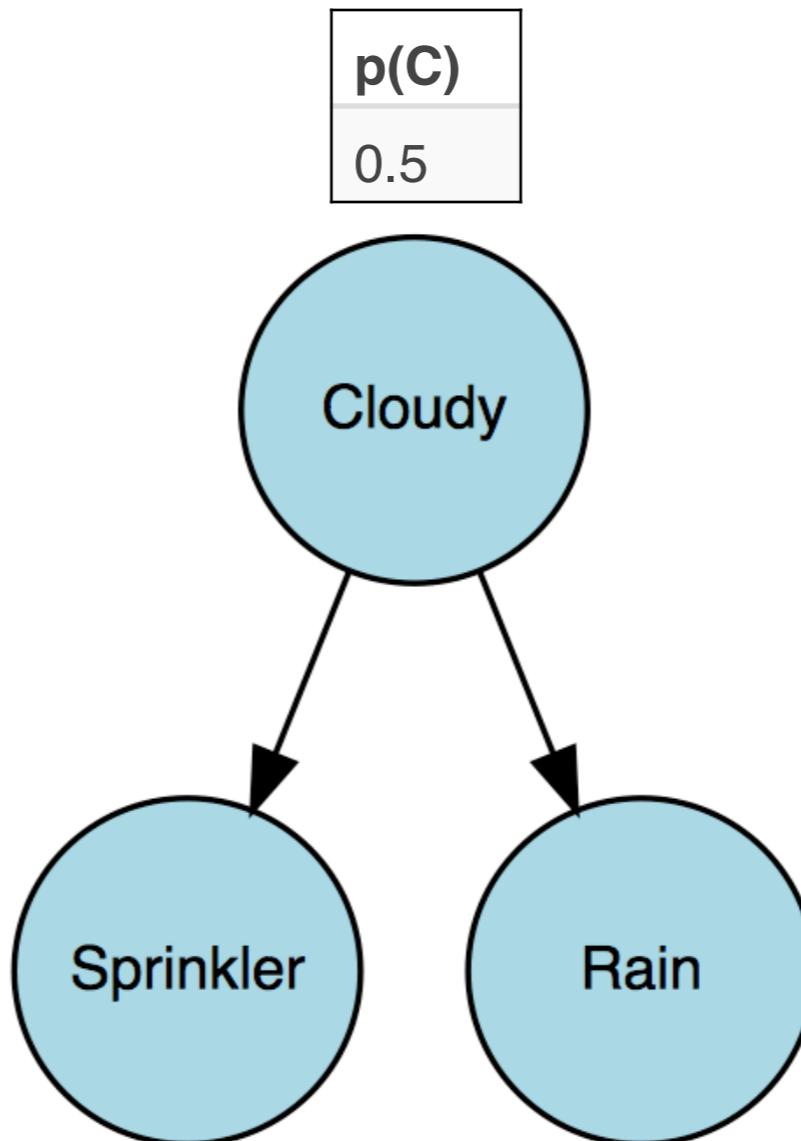
Patterns of inference: Common cause

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

or

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

?



C	p(S)
F	0.5
T	0.1

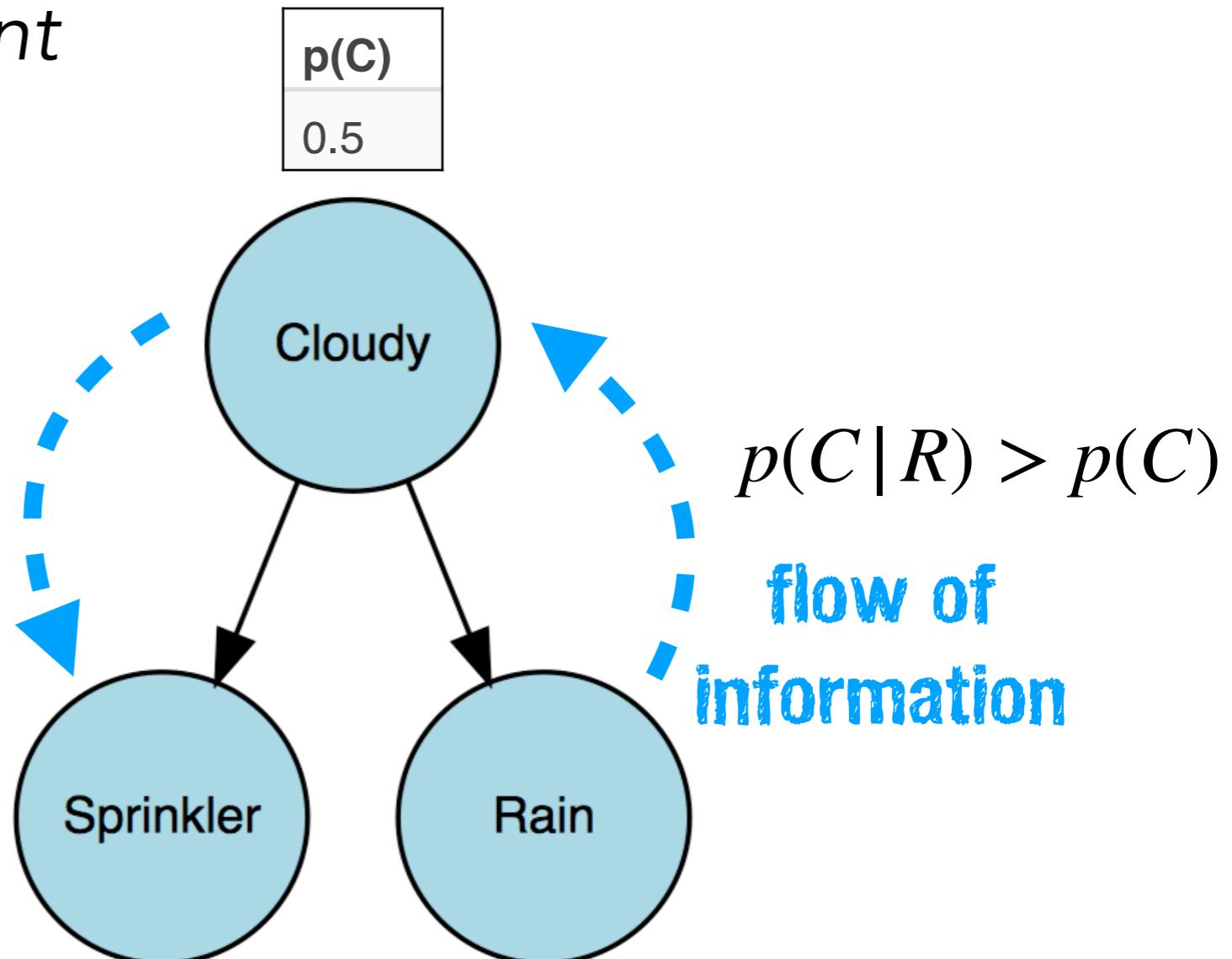
C	p(R)
F	0
T	0.3

Patterns of inference: Common cause

- effects of a common cause are *unconditionally dependent*

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

$$p(S|C) < p(S)$$



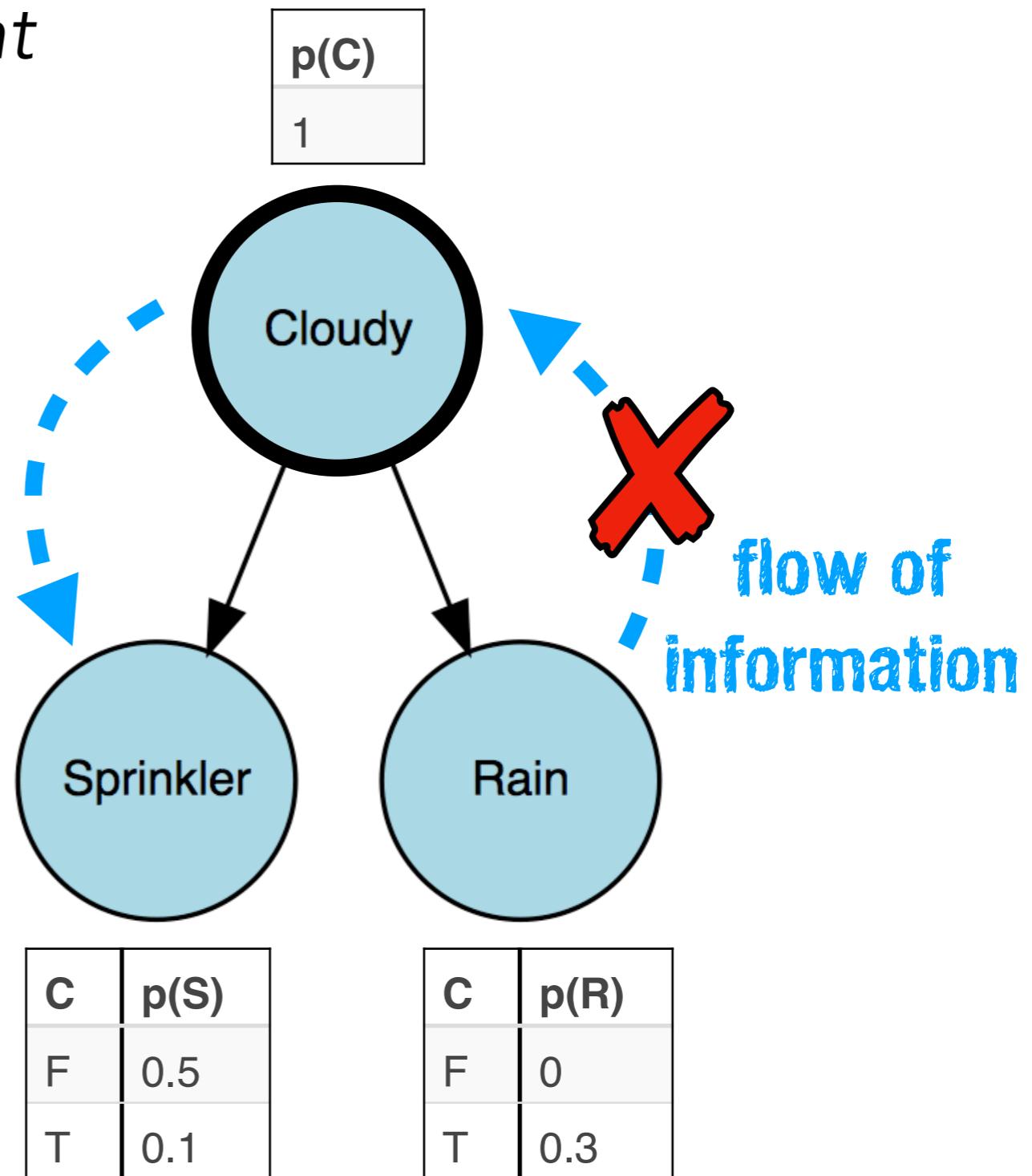
C	$p(S)$
F	0.5
T	0.1

C	$p(R)$
F	0
T	0.3

Patterns of inference: Common cause

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

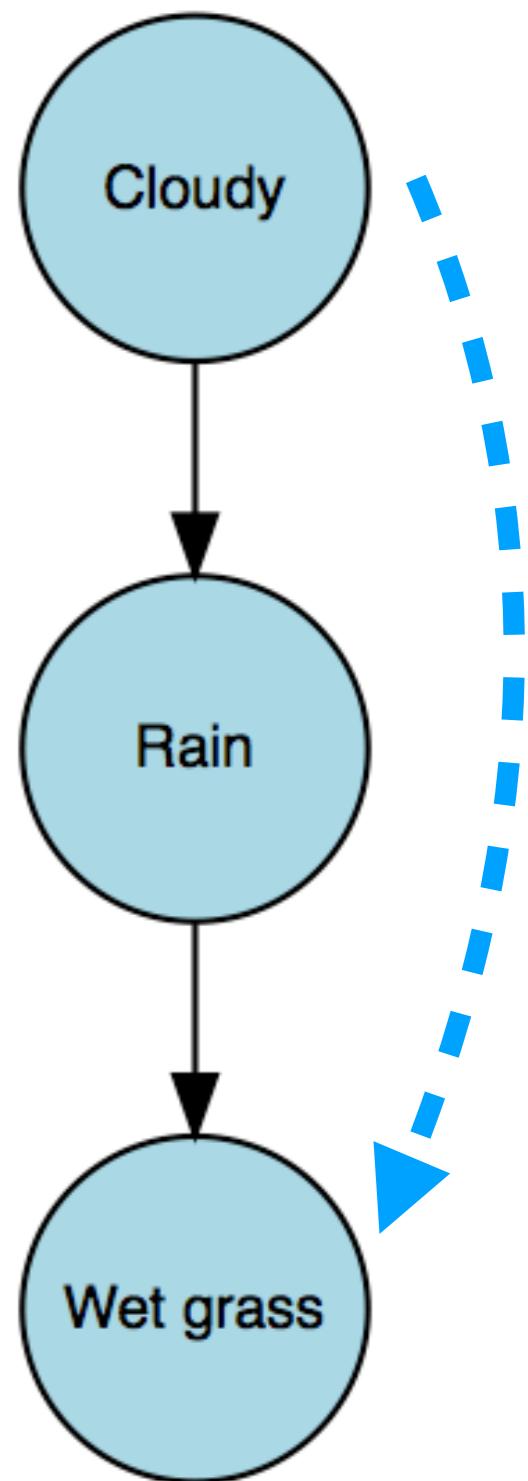
$$p(S | R, C) = p(S | C)$$



Patterns of inference: Causal chain

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

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

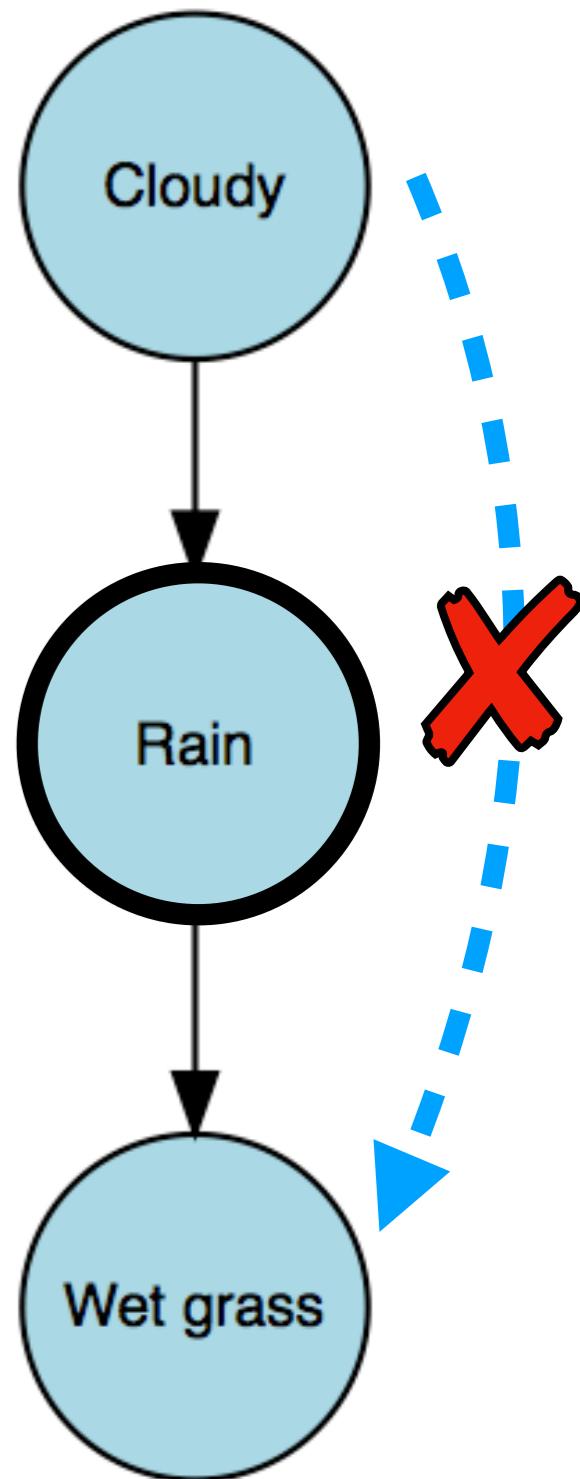


Patterns of inference: Causal chain

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

$$p(W | C, R) = p(W | R)$$

screening off

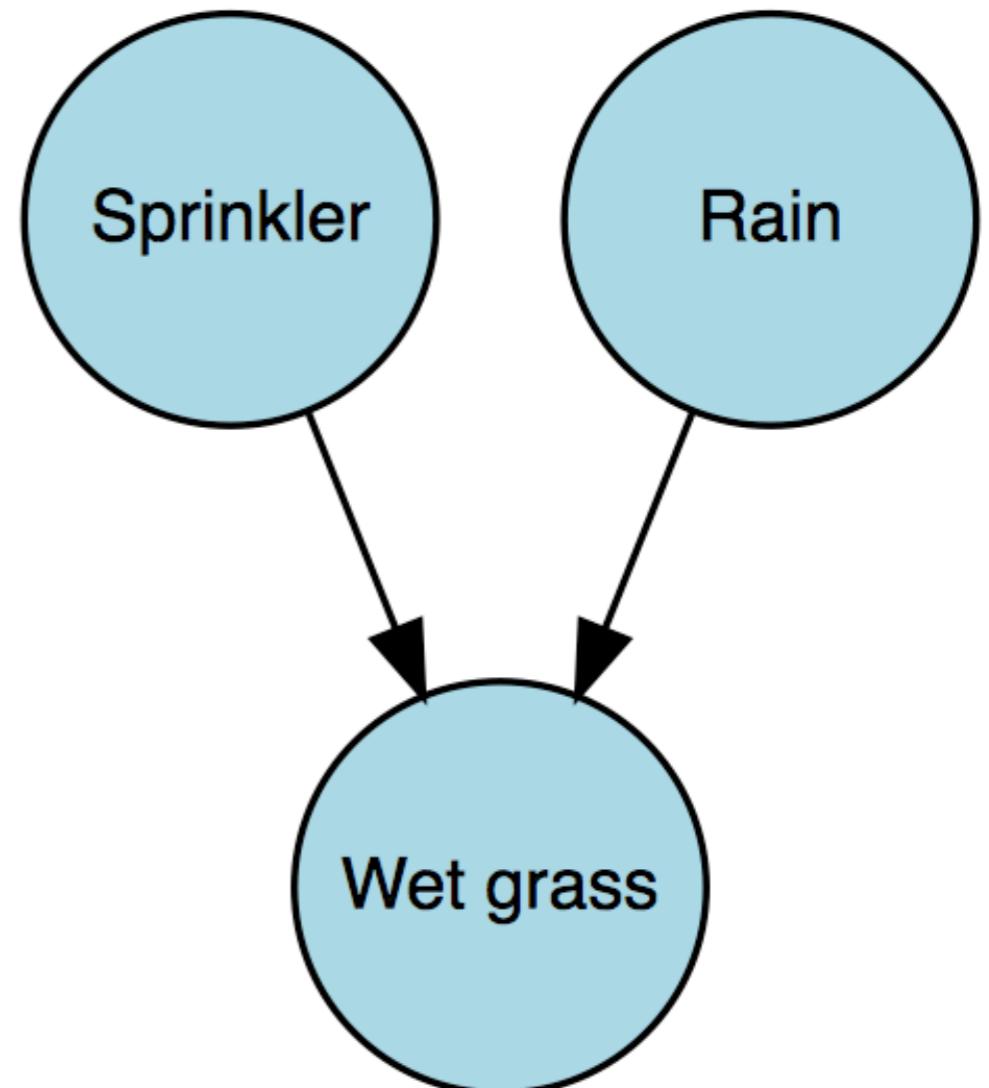


Patterns of inference: Common effect

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

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

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

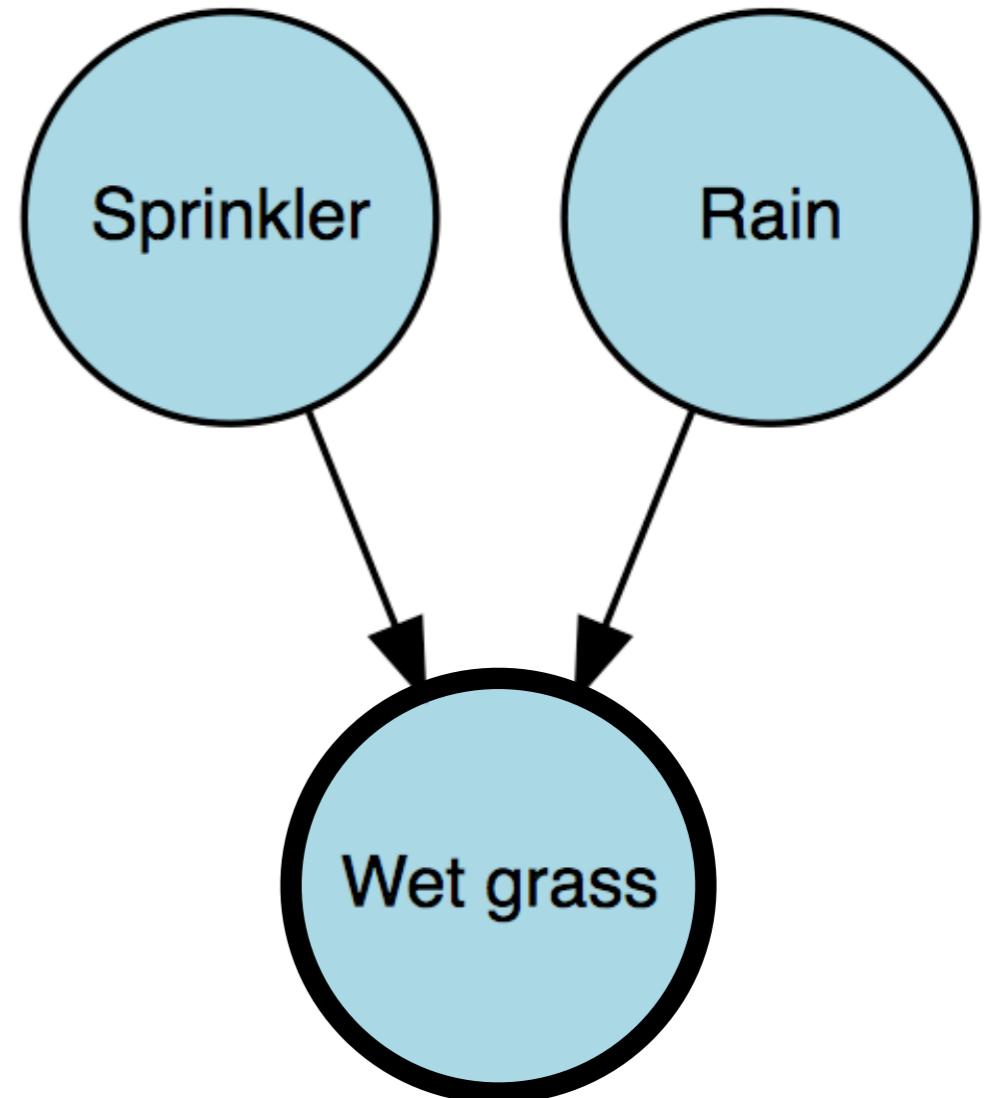


Patterns of inference: Common effect

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

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

explaining away



- intuitively: both causes compete to explain the effect

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.

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**

When should I control for variables?

How can I tell whether two variables are independent?

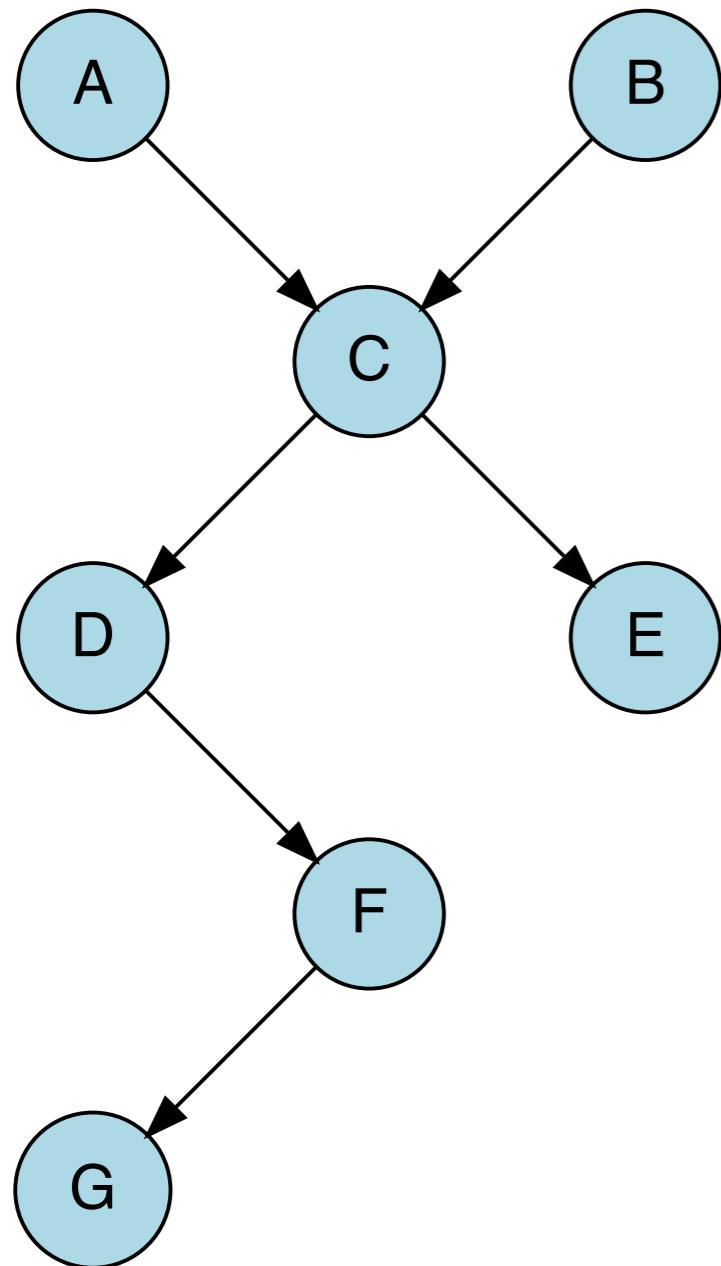
Recipe for independence

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

When should I control for variables?

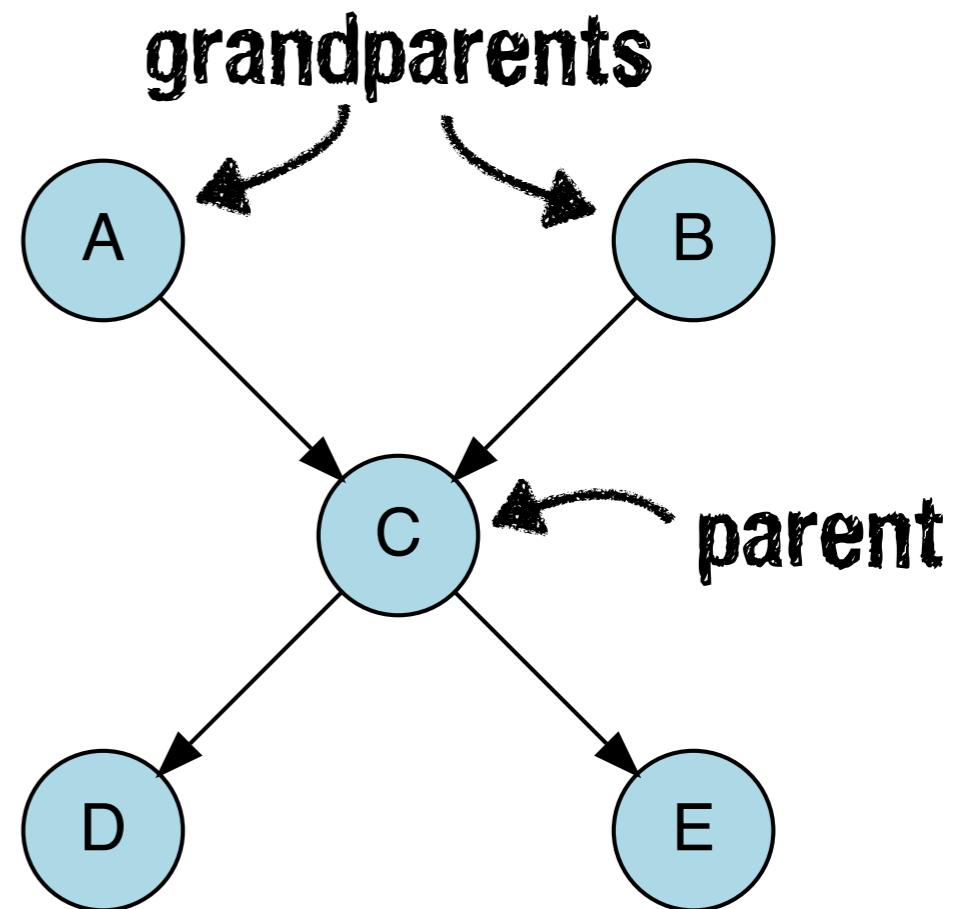
Are D and E independent?

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



1. Draw the ancestral graph

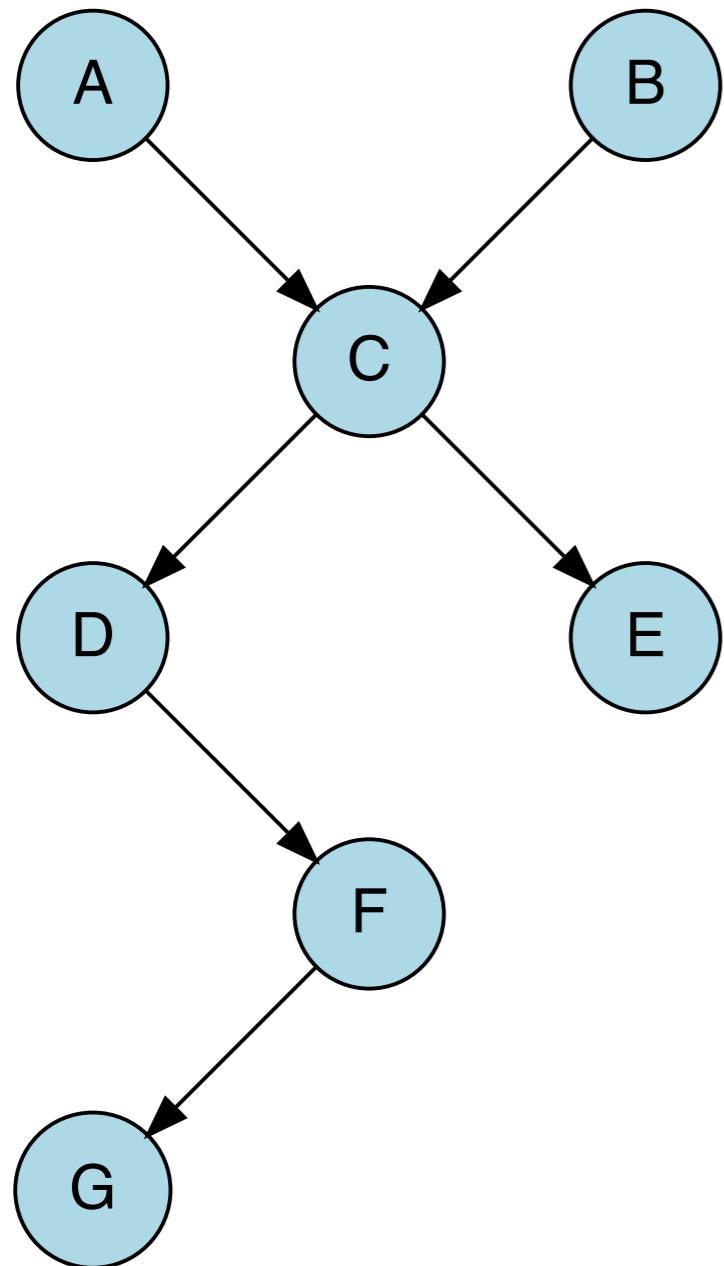
Construct the "ancestral graph" of all variables mentioned in the probability expression. This is a reduced version of the original net, consisting only of the variables mentioned and all of their ancestors (parents, parents' parents, etc.)



When should I control for variables?

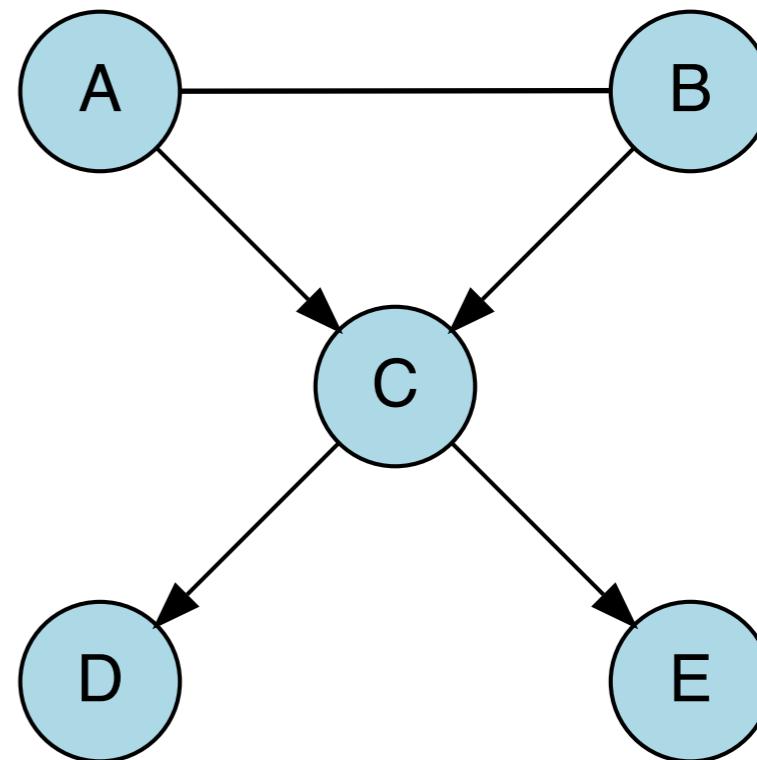
Are D and E independent?

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



2. "Moralize" the graph
let's get married!

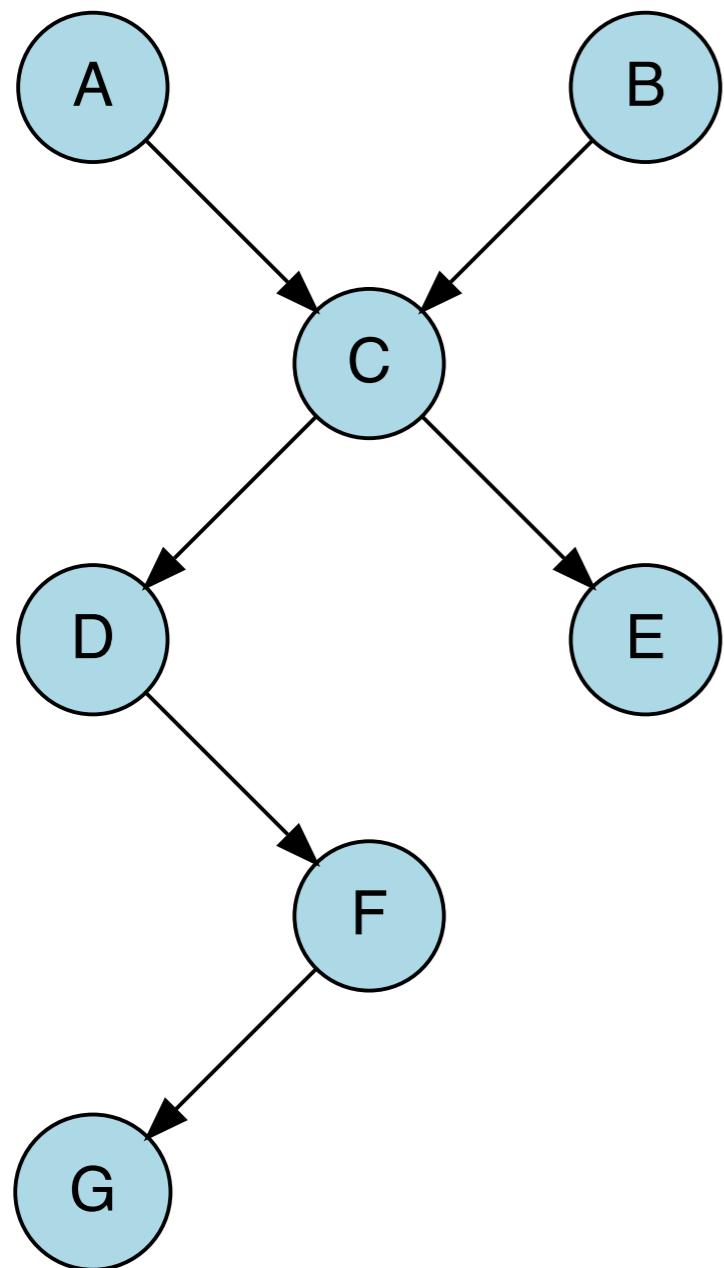
For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)



When should I control for variables?

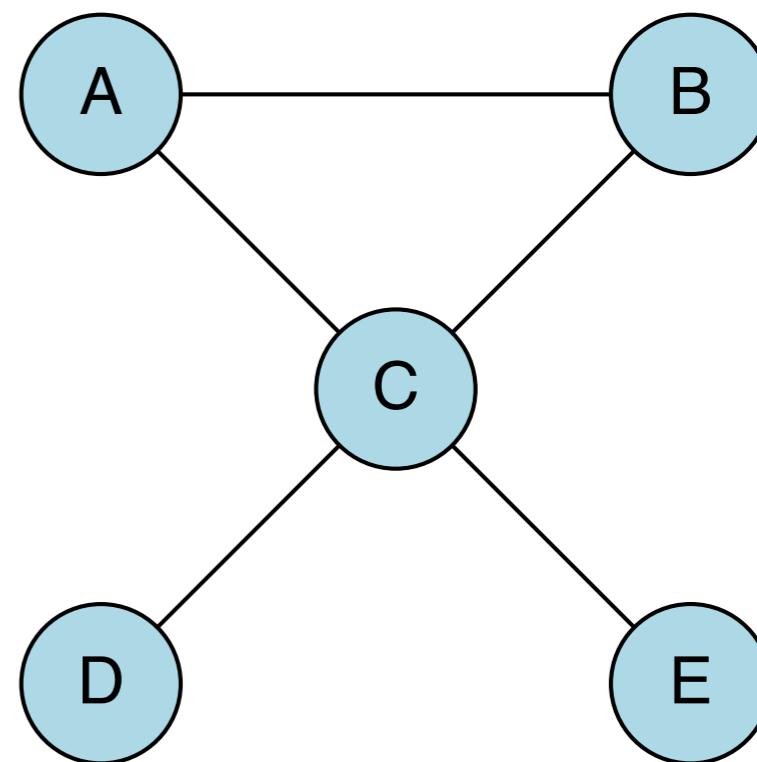
Are D and E independent?

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



3. "Disorient" the graph

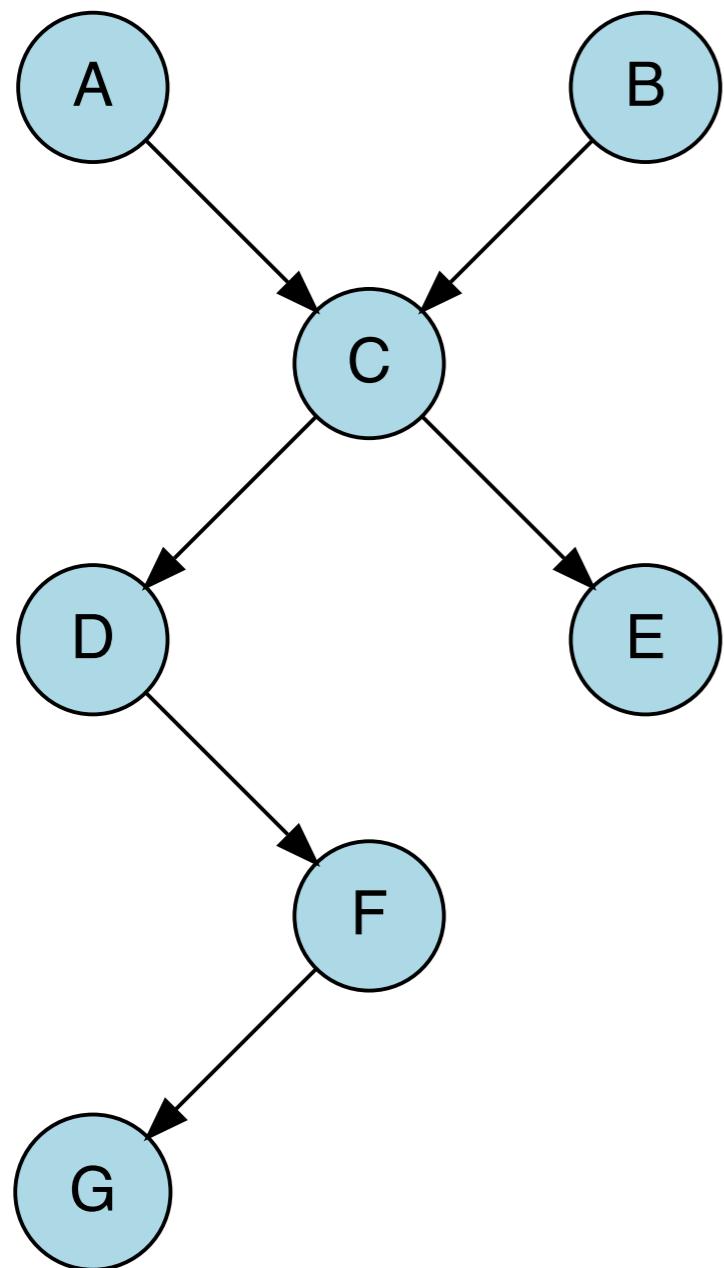
Replace arrows with lines



When should I control for variables?

Are D and E independent?

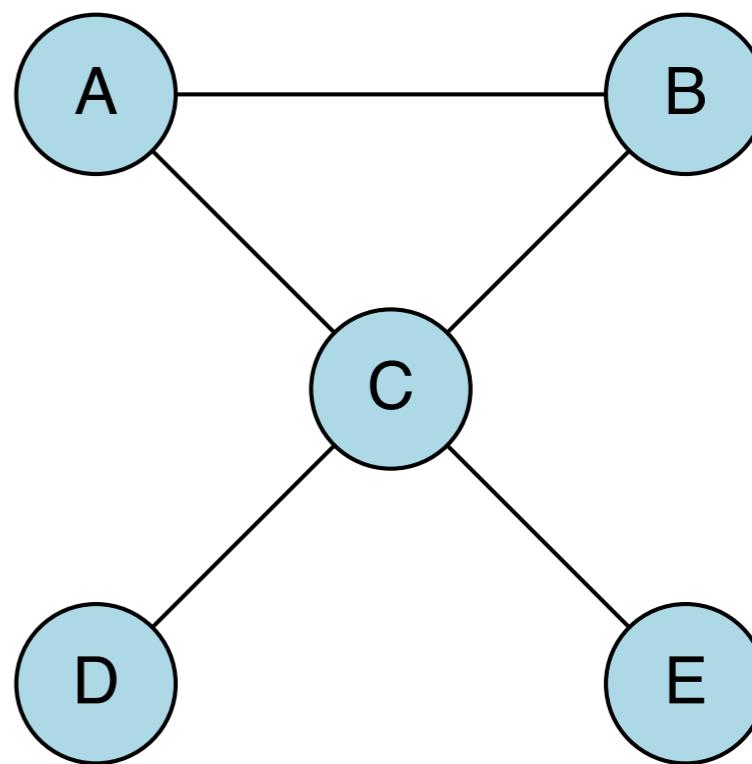
$$p(D | E) = p(D) ?$$



4. Delete the givens

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

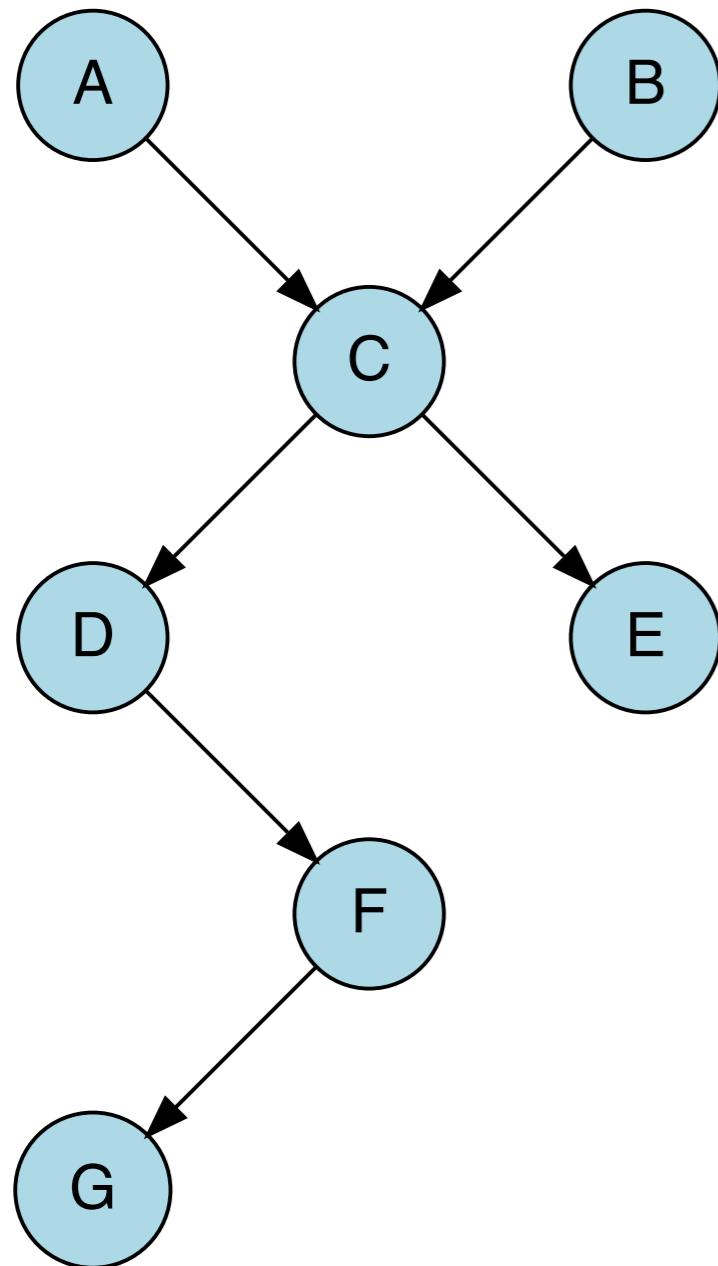
**we didn't condition on anything,
so there is nothing to delete**



When should I control for variables?

Are D and E independent?

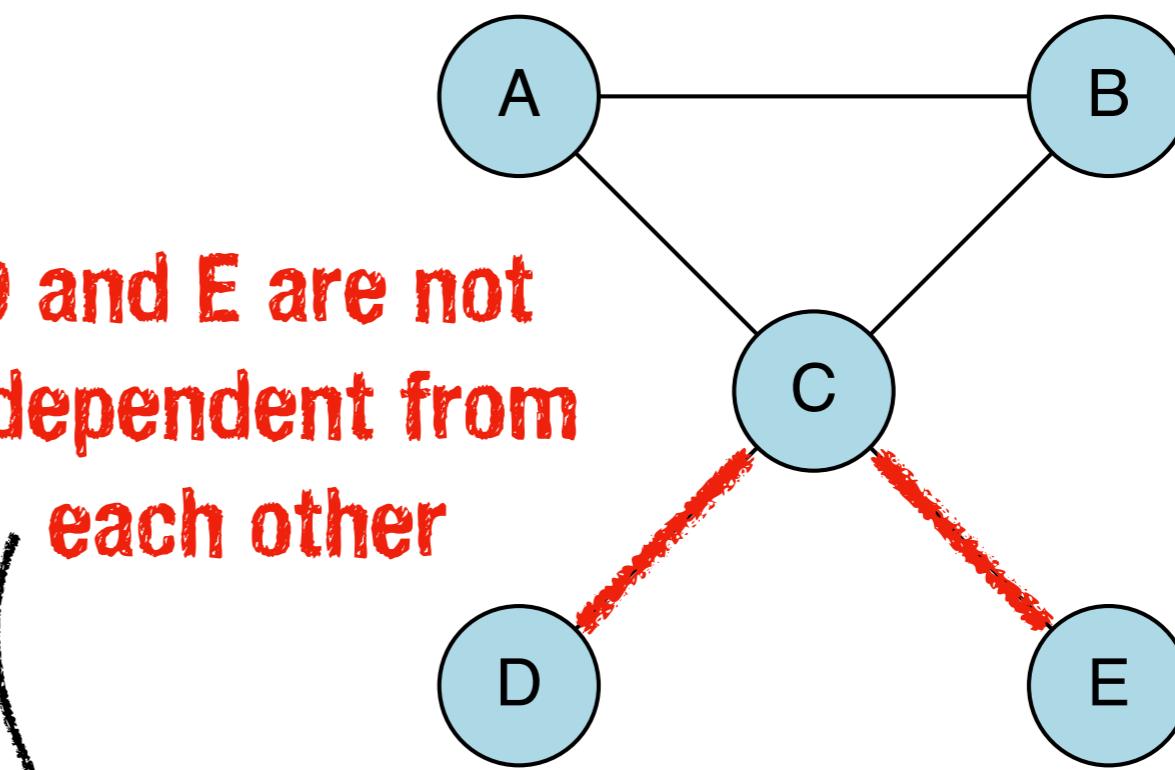
$$p(D | E) = p(D) ?$$



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 not independent from each other

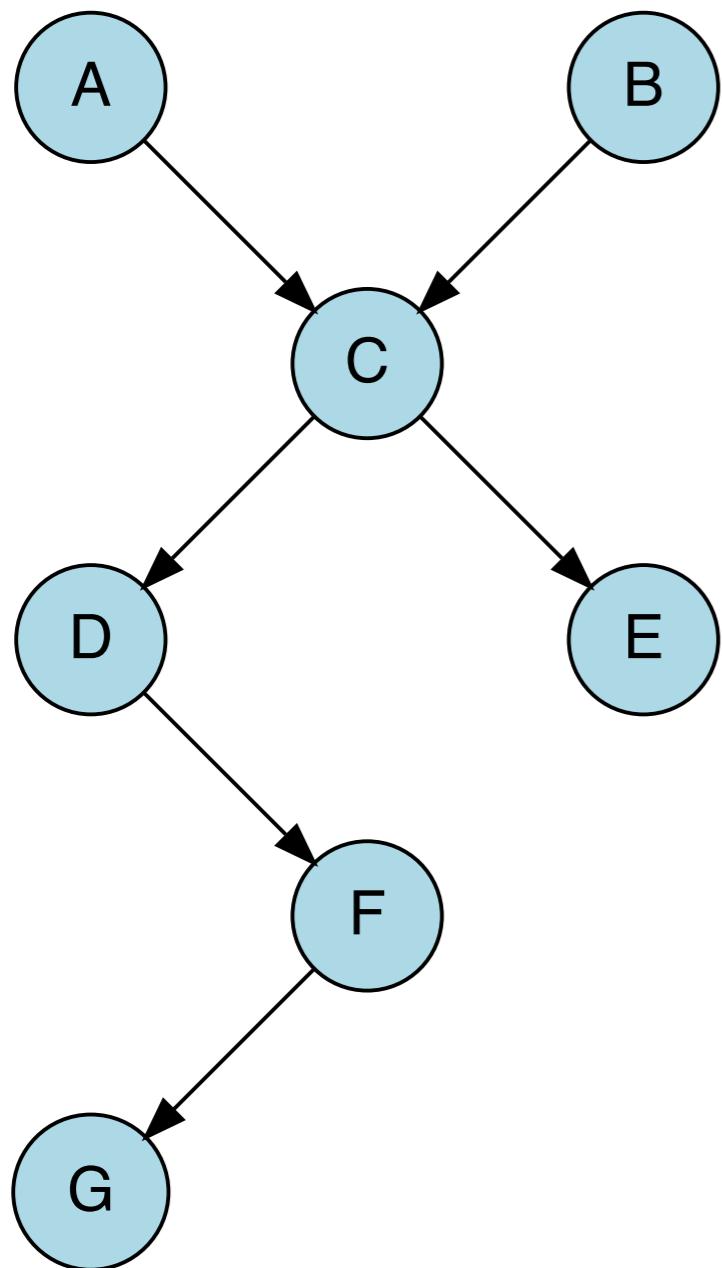


they are connected via at least one path

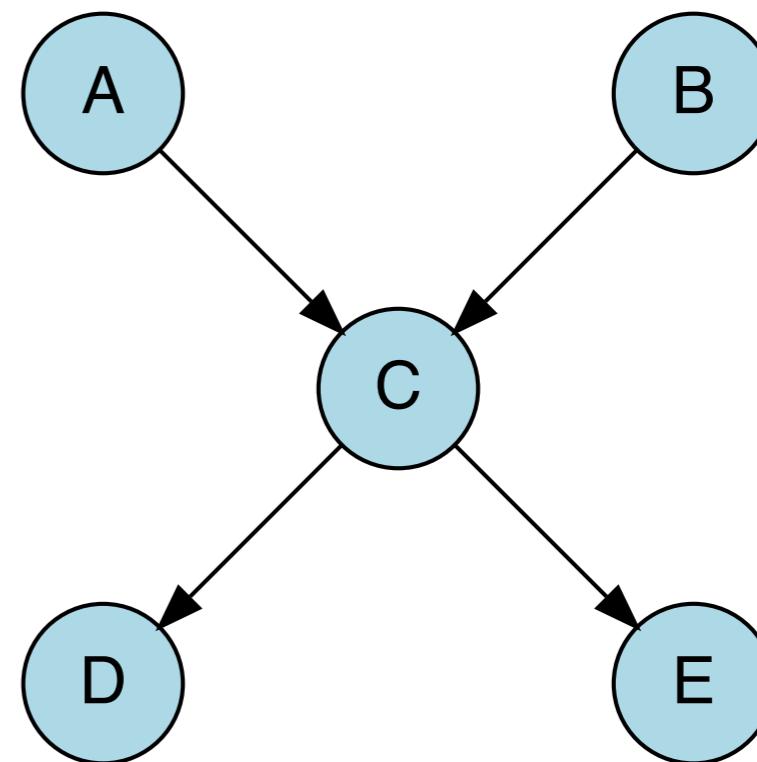
When should I control for variables?

Are D and E independent, given C? 1. Draw the ancestral graph

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



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.)

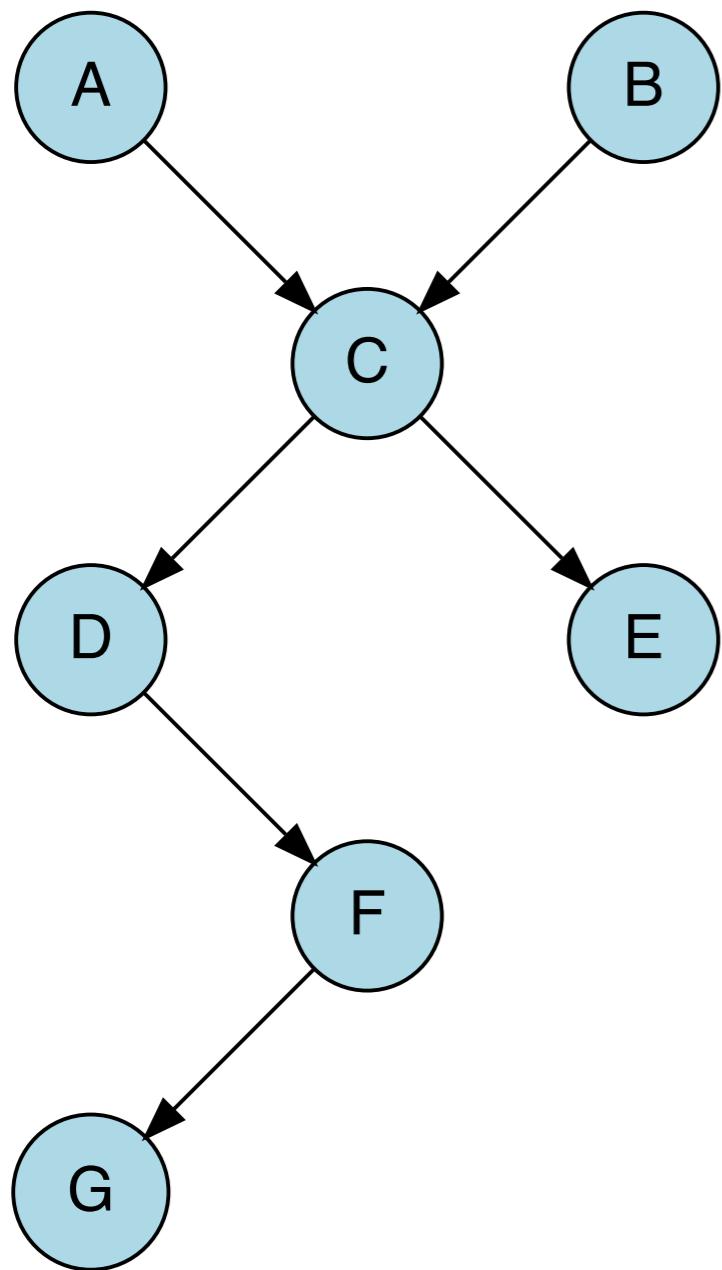


When should I control for variables?

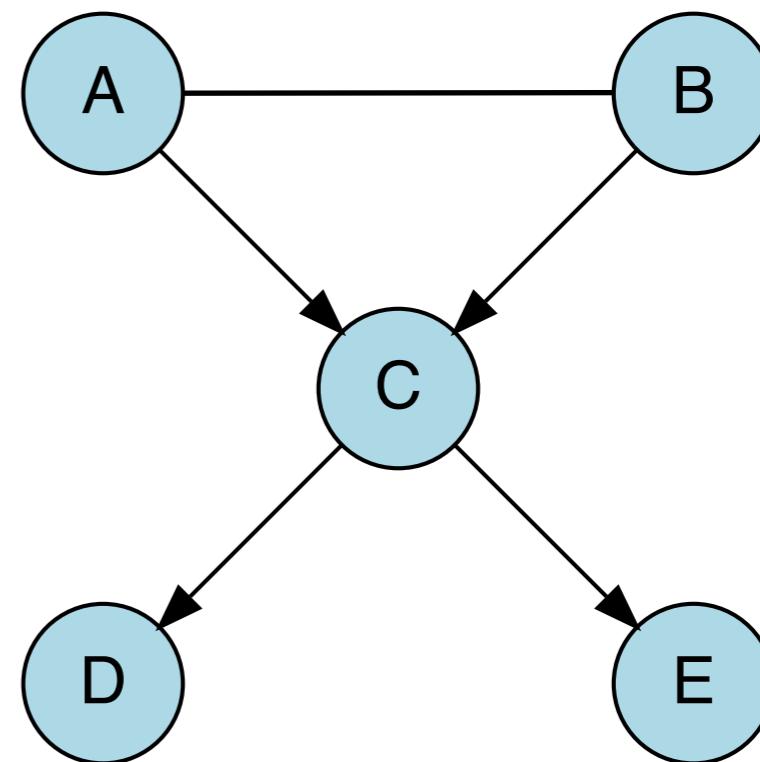
Are D and E independent, given C? 2. "Moralize" the graph

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

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.)

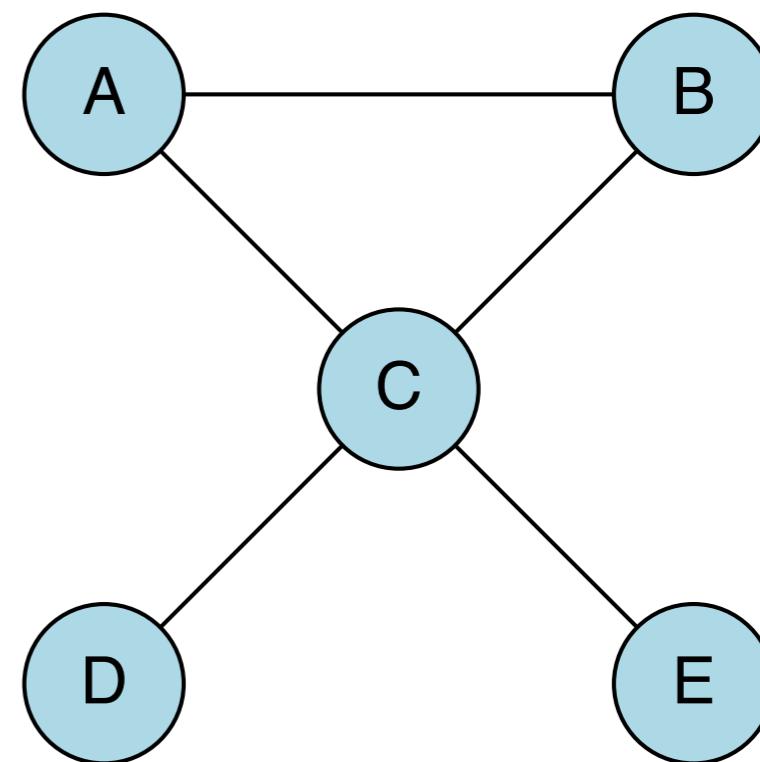
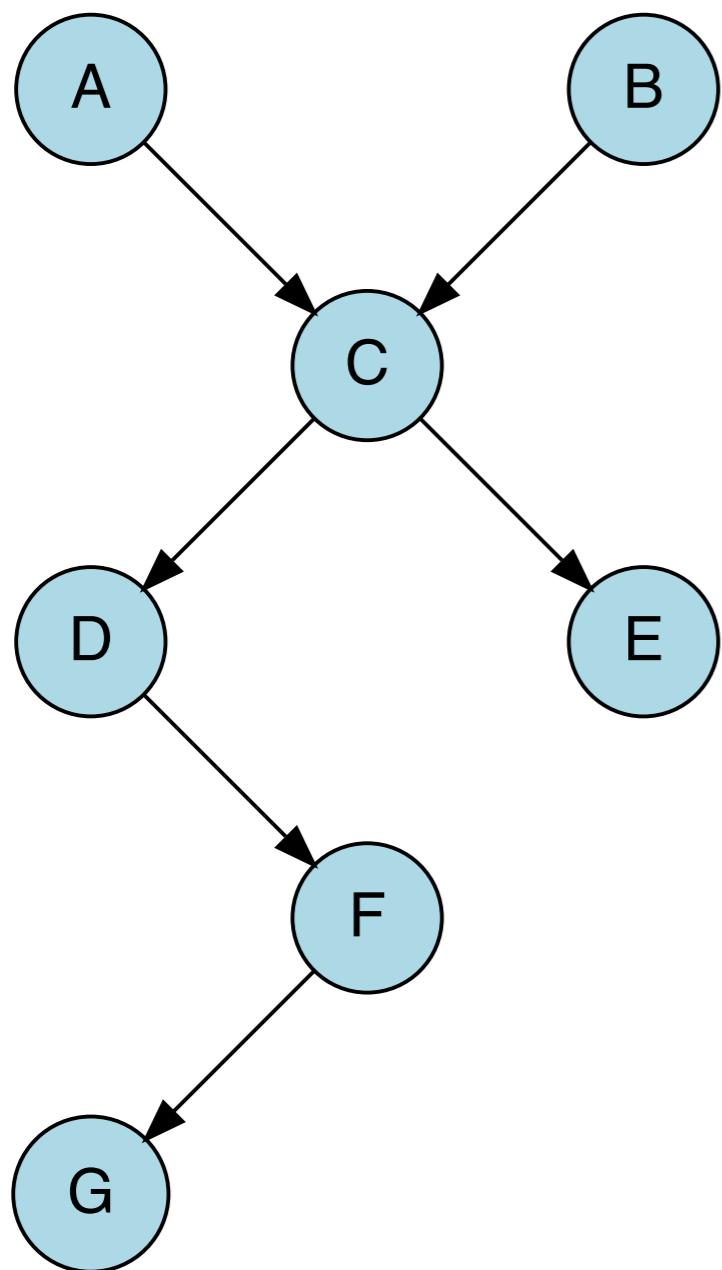


When should I control for variables?

Are D and E independent, given C? 3. "Disorient" the graph

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

Replace arrows with lines



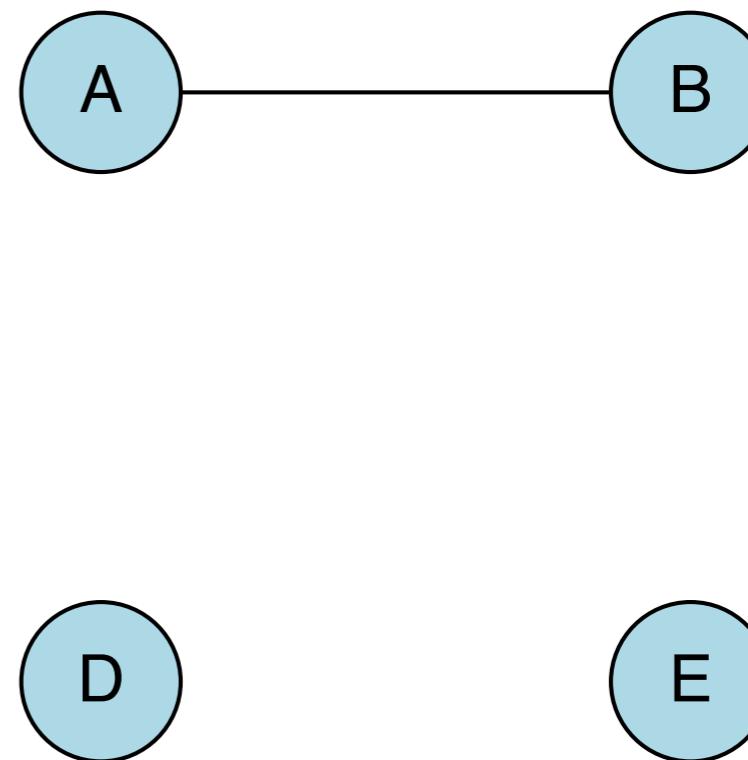
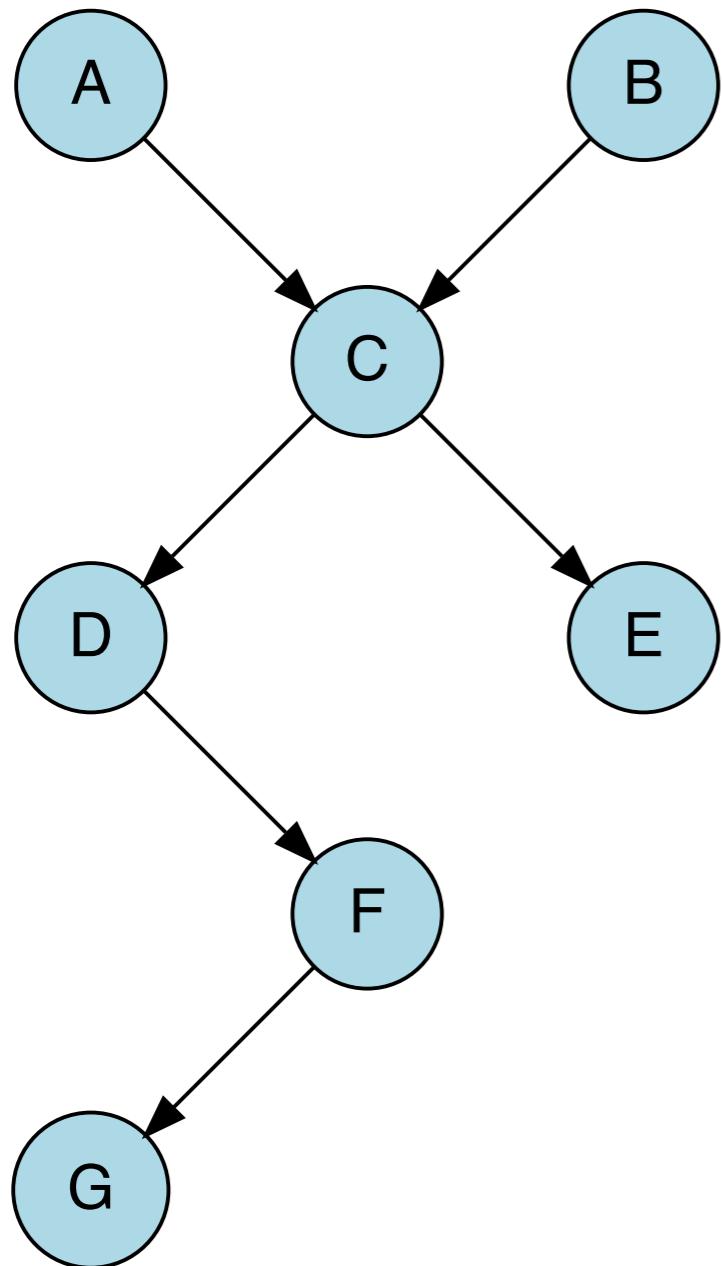
When should I control for variables?

Are D and E independent, given C? 4. Delete the givens

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

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

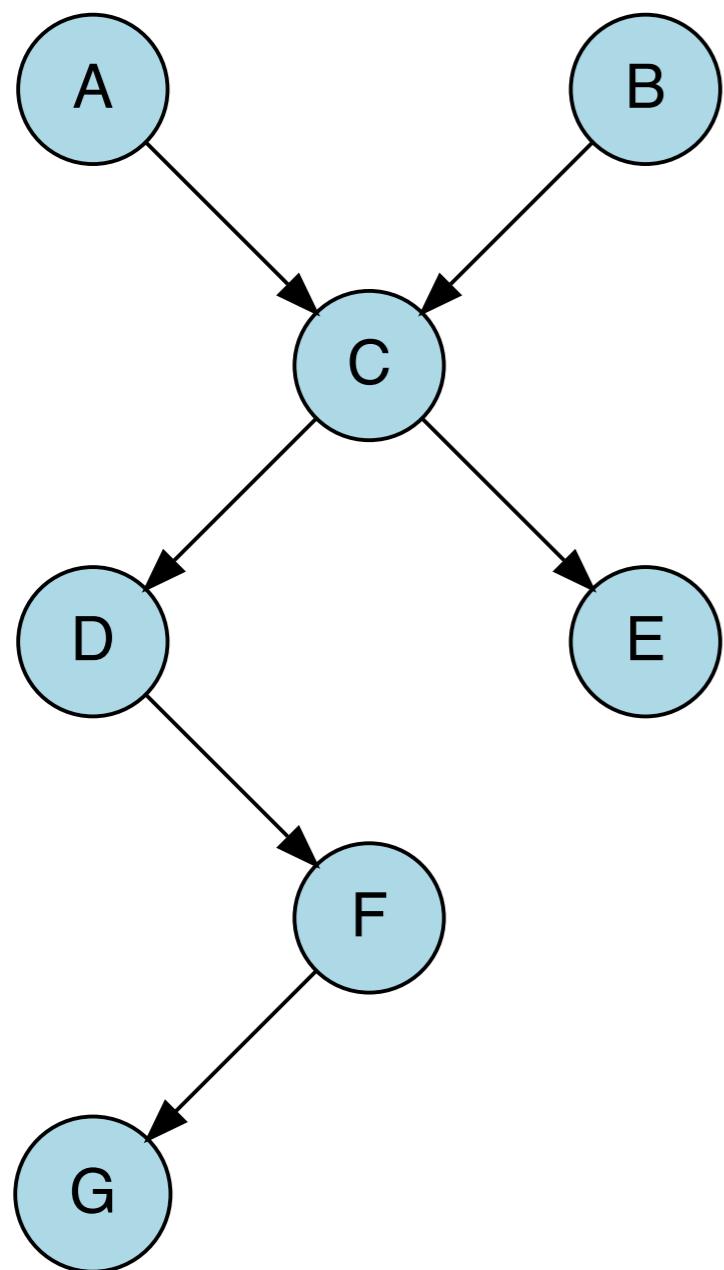
we conditioned on C!



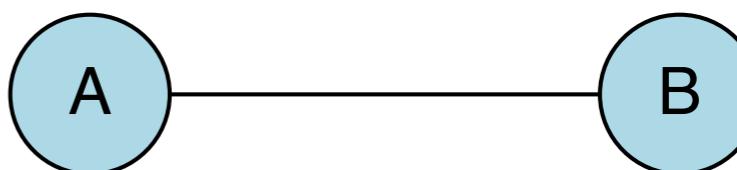
When should I control for variables?

Are D and E independent, given C? 5. Read answer off the graph

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



- 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



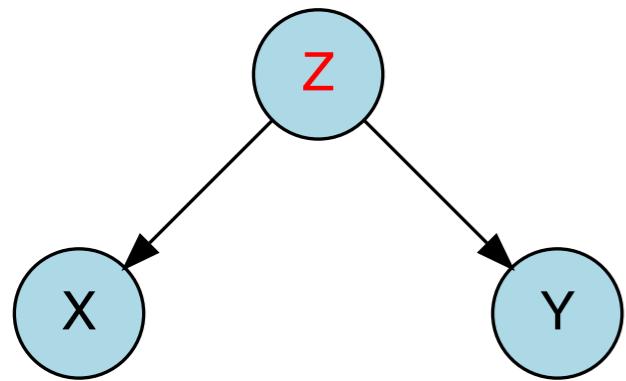
they aren't connected via a path

So what?

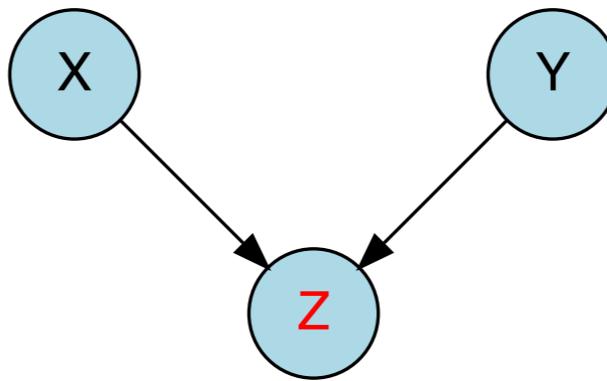


Patterns of inference

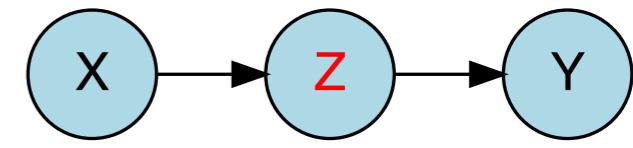
We want to estimate the (causal) relationship between X and Y



common cause



common effect



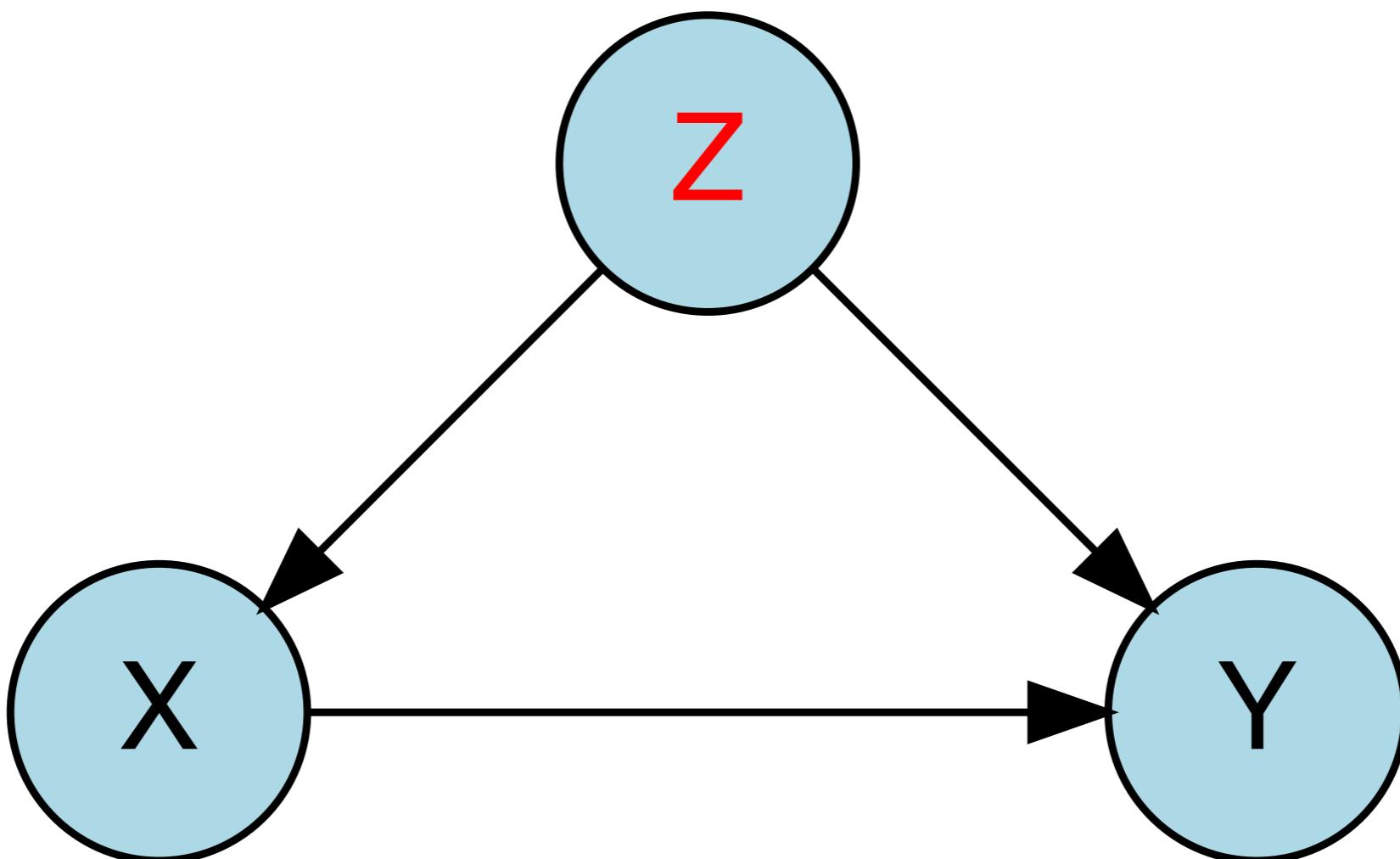
causal chain

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?

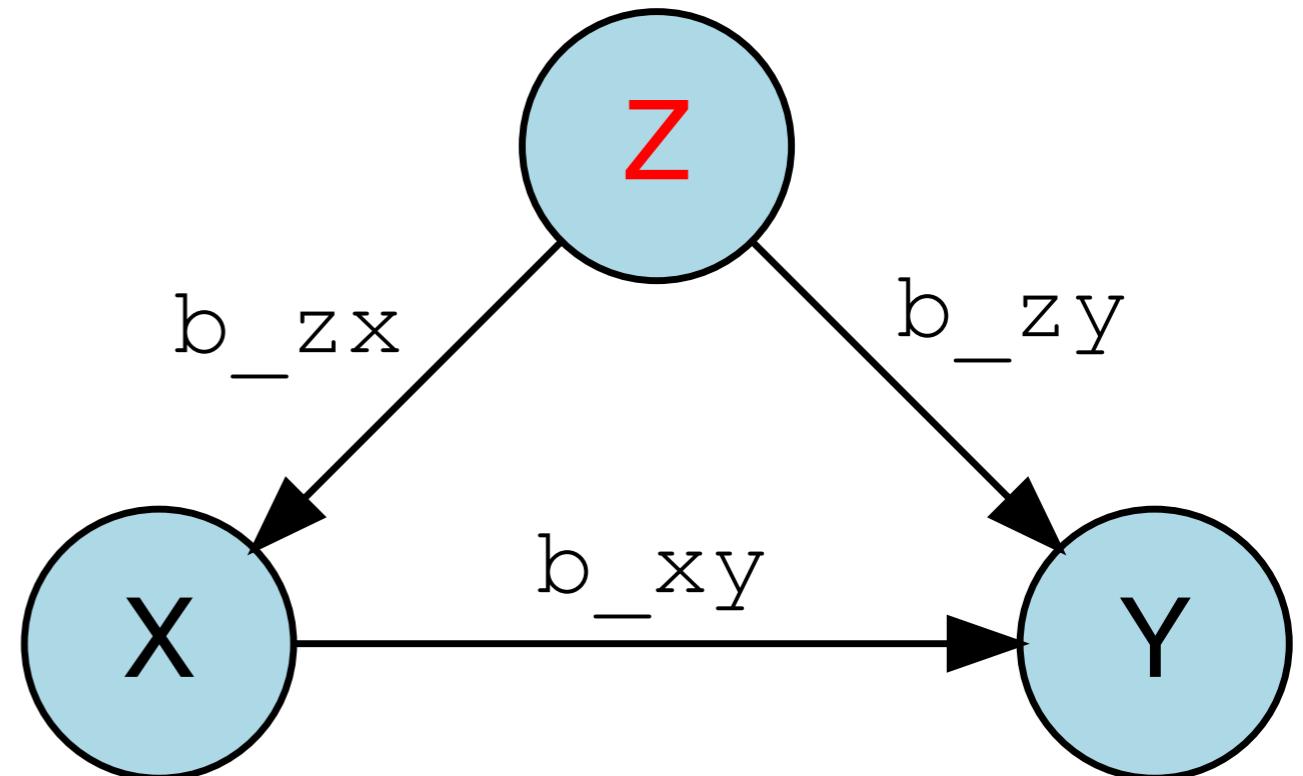
I want to estimate the effect that X has on Y



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

When should I control for variables?

```
1 set.seed(1)
2
3 n = 1000
4 b_zx = 2
5 b_xy = 2
6 b_zy = 2
7 sd = 1
8
9 fun_error = function(n, sd) {
10   rnorm(n = n,
11         mean = 0,
12         sd = sd)
13 }
14
15 df = tibble(z = fun_error(n, sd),
16               x = b_zx * z + fun_error(n, sd),
17               y = b_zy * z + b_xy * x + fun_error(n, sd))
```



overestimating
X's effect on Y

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

```
1 # without control
2 lm(formula = y ~ x,
3     data = df) %>%
4   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

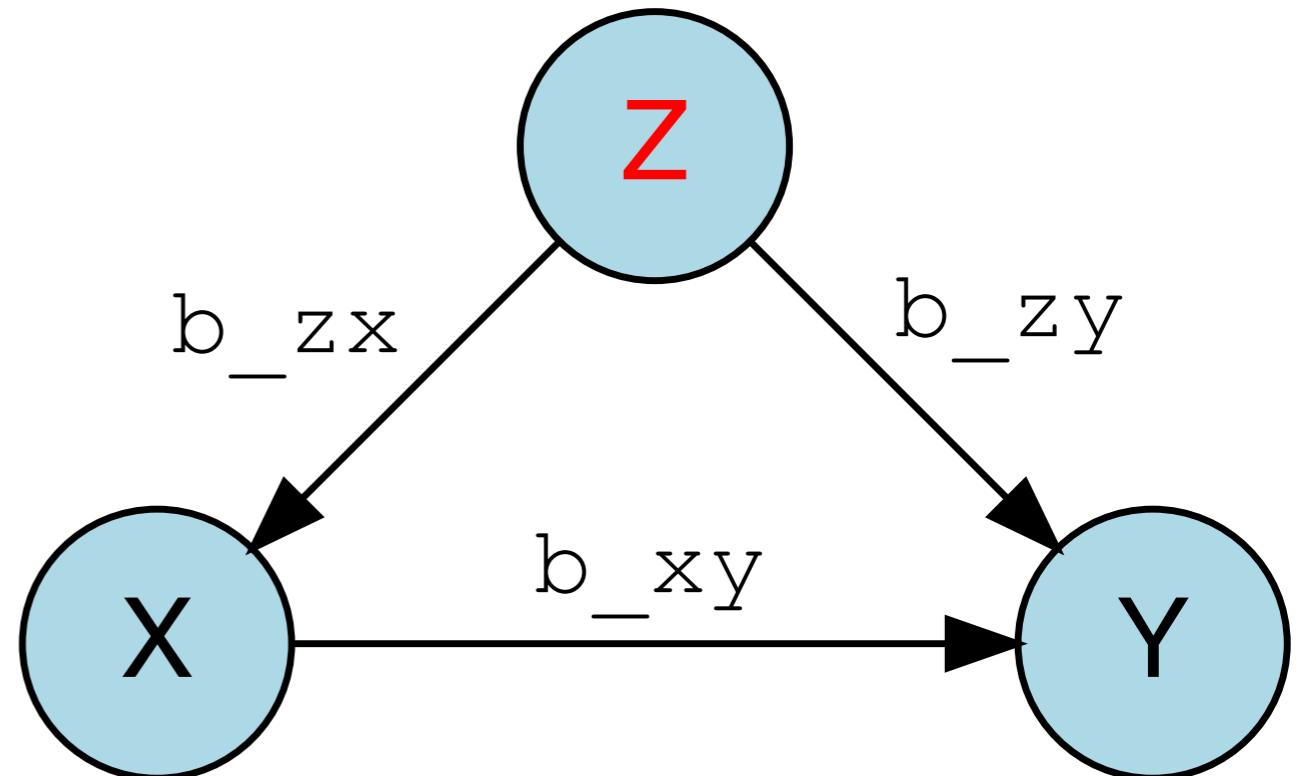
Residual standard error: 1.388 on 998 degrees of freedom
Multiple R-squared:  0.9571,    Adjusted R-squared:  0.9571 
F-statistic: 2.227e+04 on 1 and 998 DF,  p-value: < 2.2e-16
```

When should I control for variables?

```
1 set.seed(1)
2
3 n = 1000
4 b_zx = 2
5 b_xy = 2
6 b_zy = 2
7 sd = 1
8
9 fun_error = function(n, sd) {
10   rnorm(n = n,
11         mean = 0,
12         sd = sd)
13 }
14
15 df = tibble(z = fun_error(n, sd),
16               x = b_zx * z + fun_error(n, sd),
17               y = b_zy * z + b_xy * x + fun_error(n, sd))
```

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

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



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 ***

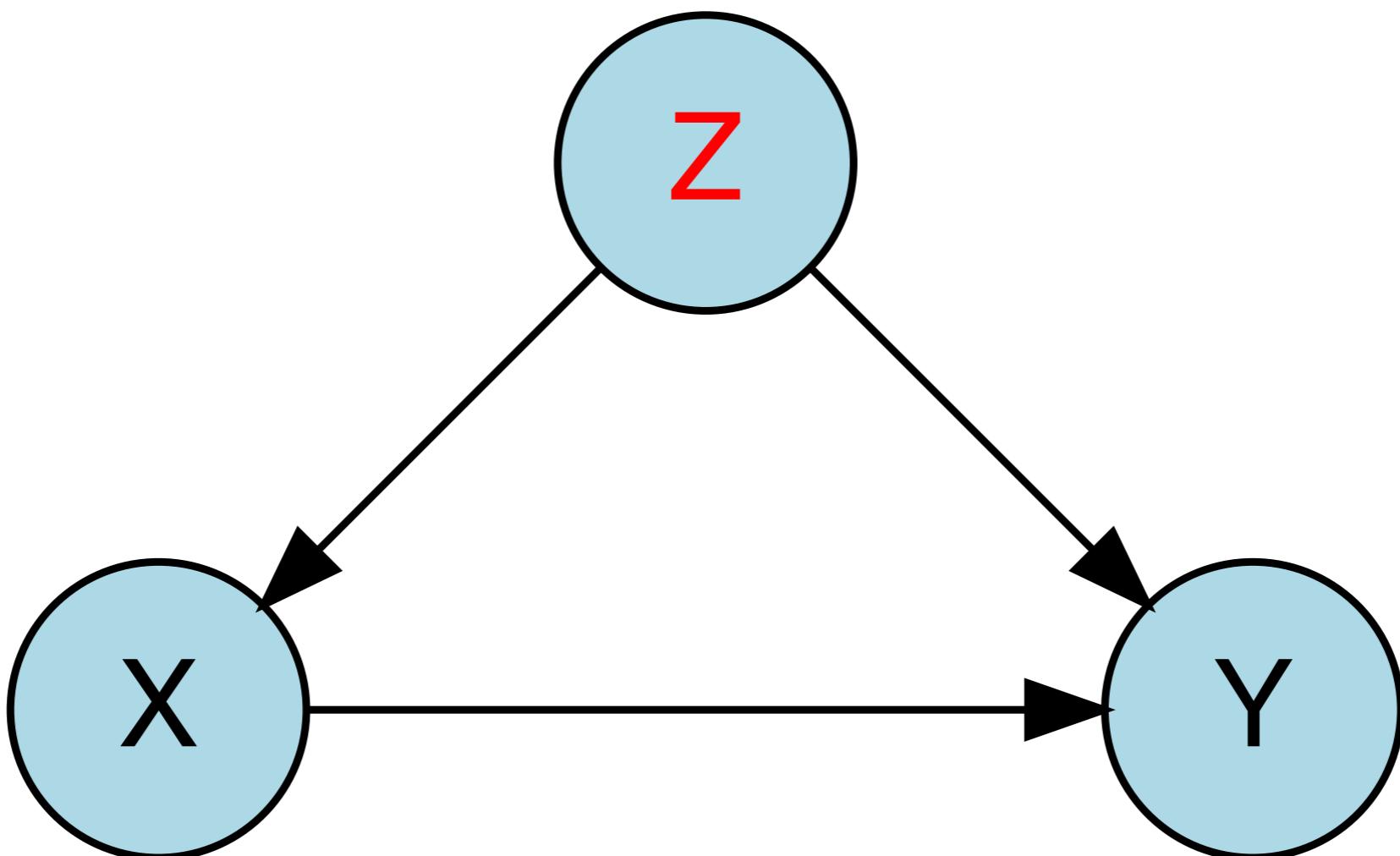
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

**accurate estimate
of X's effect on Y**

When should I control for variables?

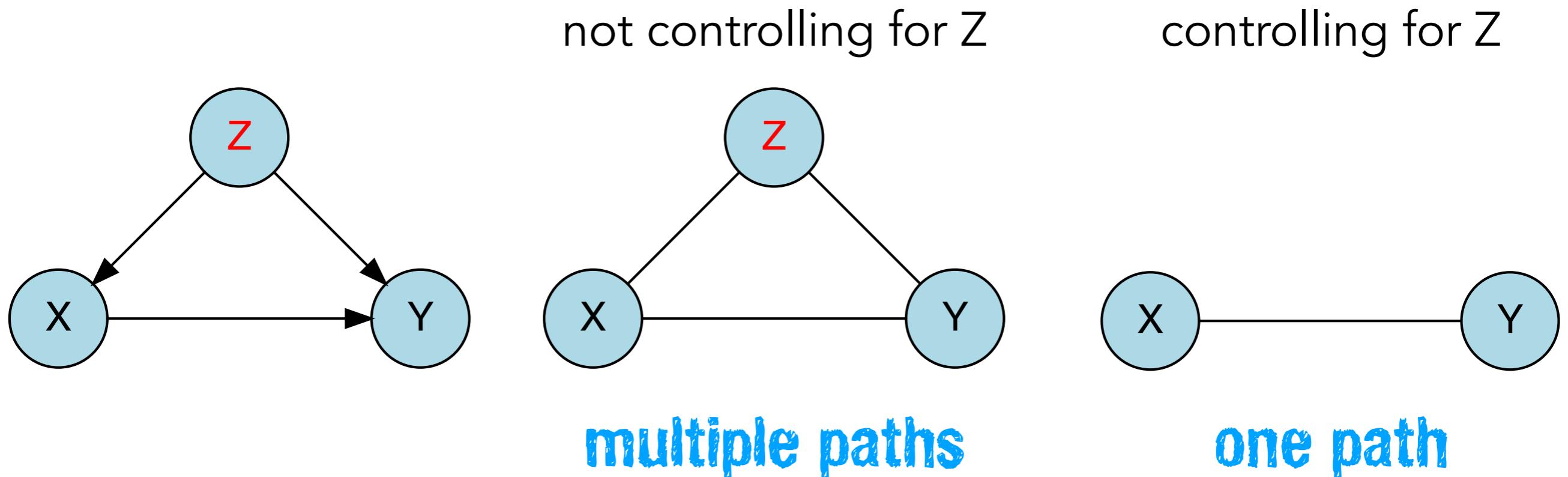
I want to estimate the effect that X has on Y



Z is a **good** control here!

When should I control for variables?

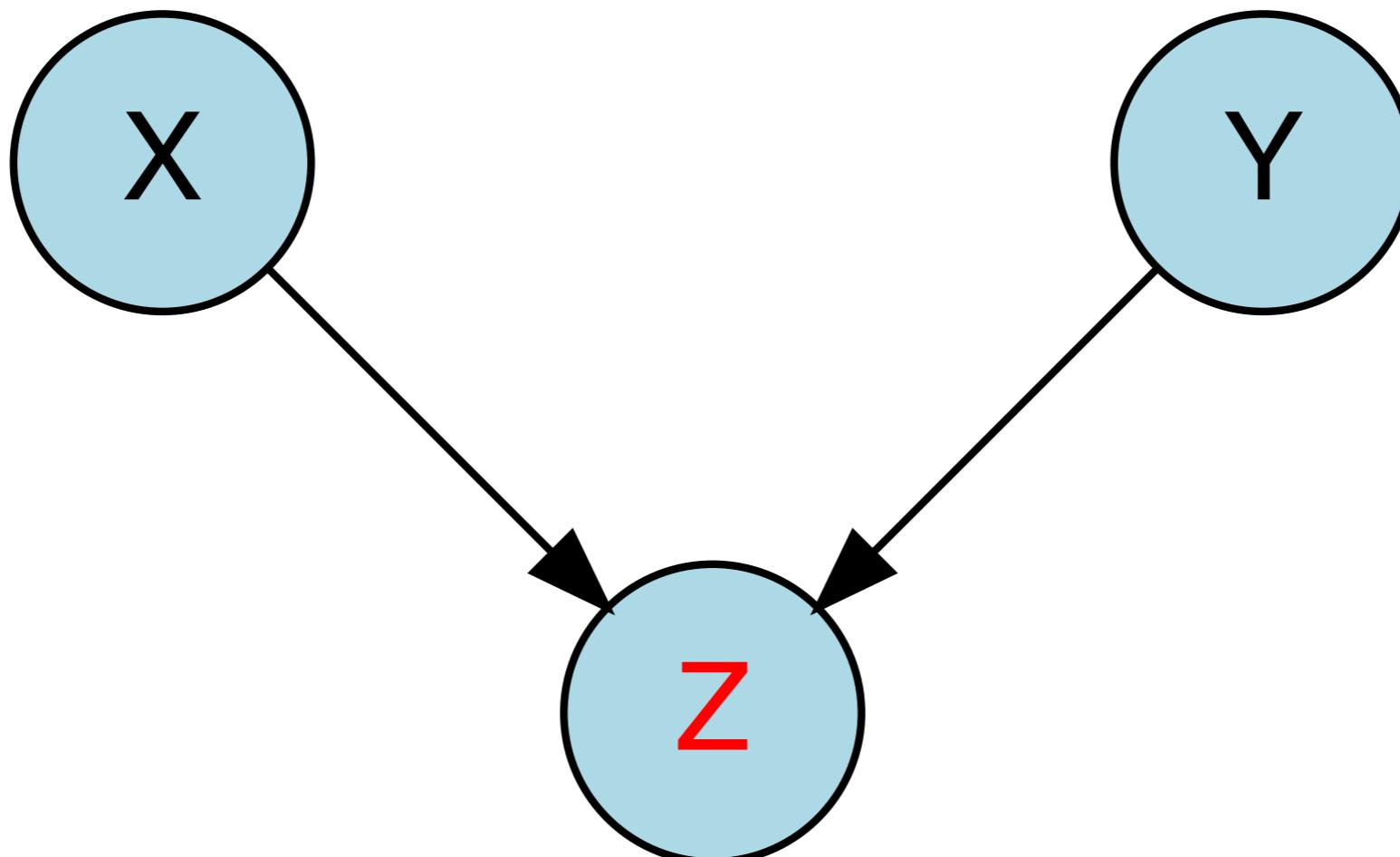
I want to estimate the effect that X has on Y



Z is a **good** control here!

When should I control for variables?

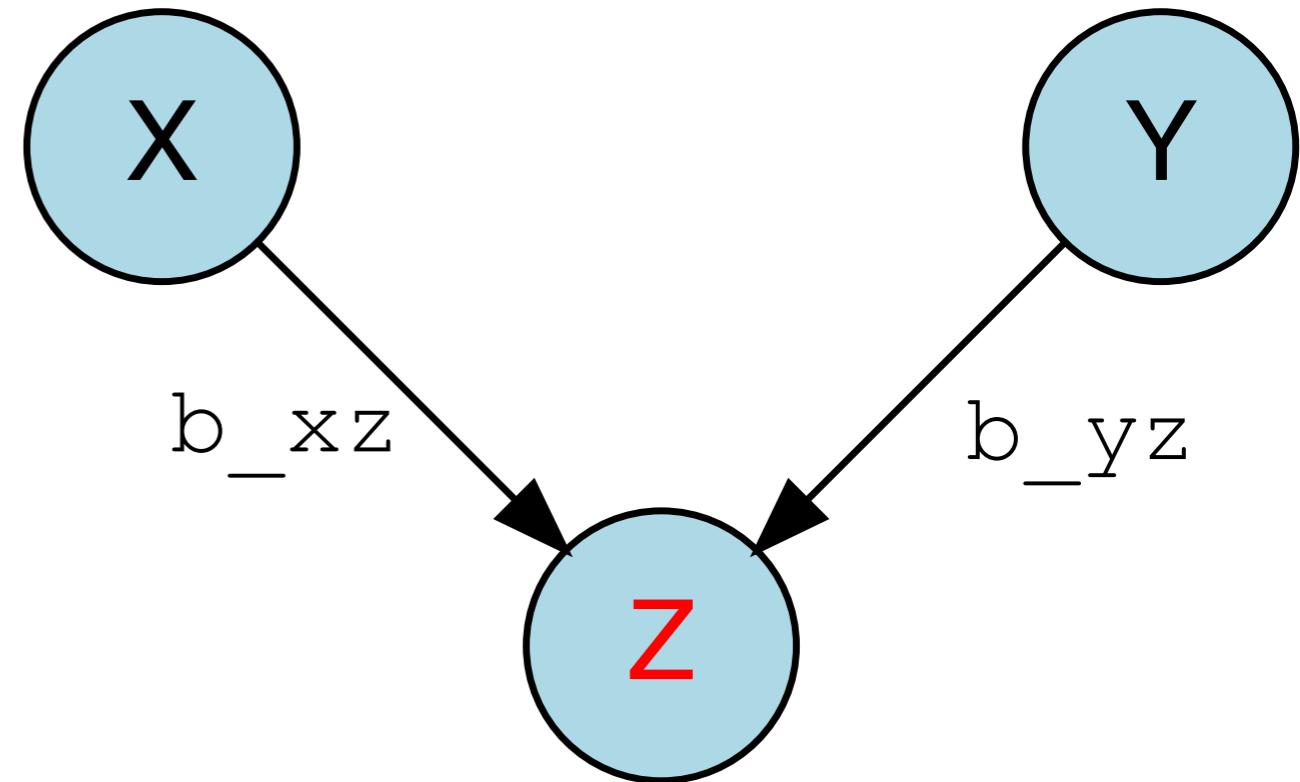
I want to estimate the effect that X has on Y



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

When should I control for variables?

```
1 set.seed(1)
2 n = 1000
3 b_xz = 2
4 b_yz = 2
5 sd = 1
6
7 fun_error = function(n, sd){
8   rnorm(n = n,
9         mean = 0,
10        sd = sd)
11 }
12
13 df = tibble(x = fun_error(n, sd),
14               y = fun_error(n, sd),
15               z = x * b_xz + y * b_yz + fun_error(n, sd))
```



accurate estimate
of X's effect on Y

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

```
1 # without control
2 lm(formula = y ~ x,
3     data = df) %>%
4   summary()
```

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

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

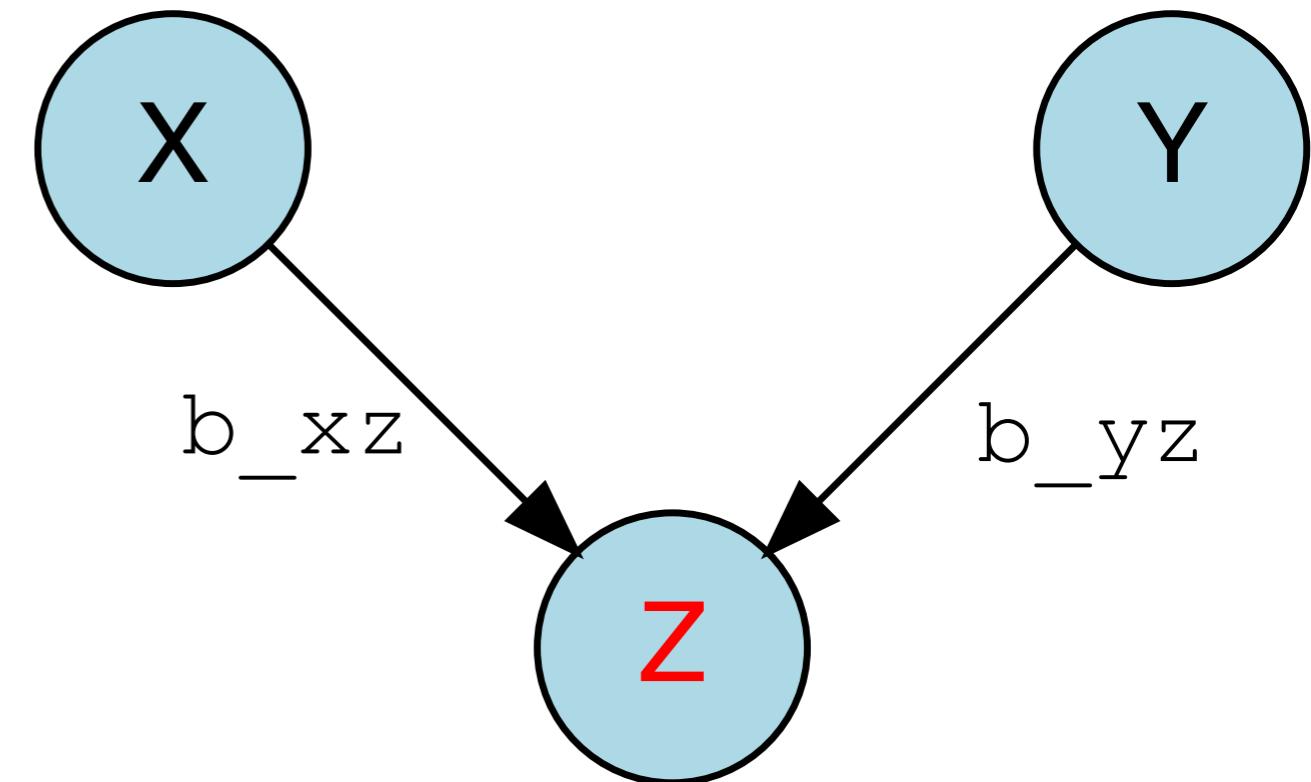
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.016187  0.032905 -0.492   0.623    
x            0.006433  0.031809  0.202   0.840    
                                                        
Residual standard error: 1.04 on 998 degrees of freedom
Multiple R-squared:  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?

```
1 set.seed(1)
2 n = 1000
3 b_xz = 2
4 b_yz = 2
5 sd = 1
6
7 fun_error = function(n, sd){
8   rnorm(n = n,
9         mean = 0,
10        sd = sd)
11 }
12
13 df = tibble(x = fun_error(n, sd),
14               y = fun_error(n, sd),
15               z = x * b_xz + y * b_yz + fun_error(n, sd))
```

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

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



inaccurate
estimate of X's
effect on Y

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

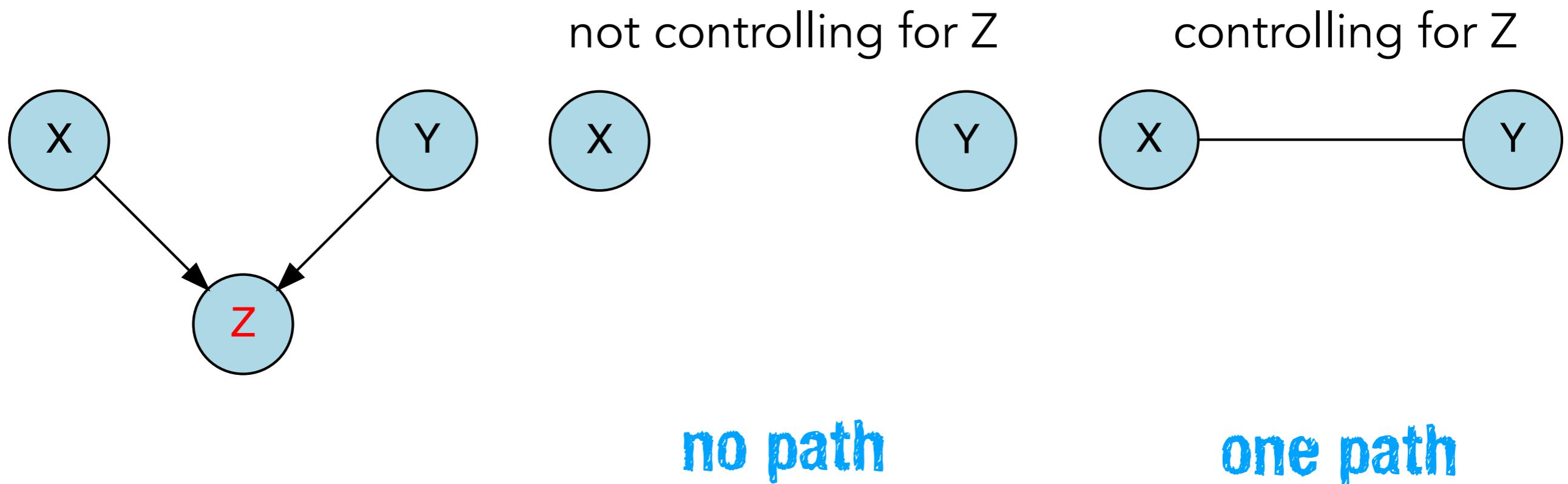
Residuals:
    Min      1Q  Median      3Q     Max 
-1.35547 -0.30016  0.00298  0.31119  1.73408 

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



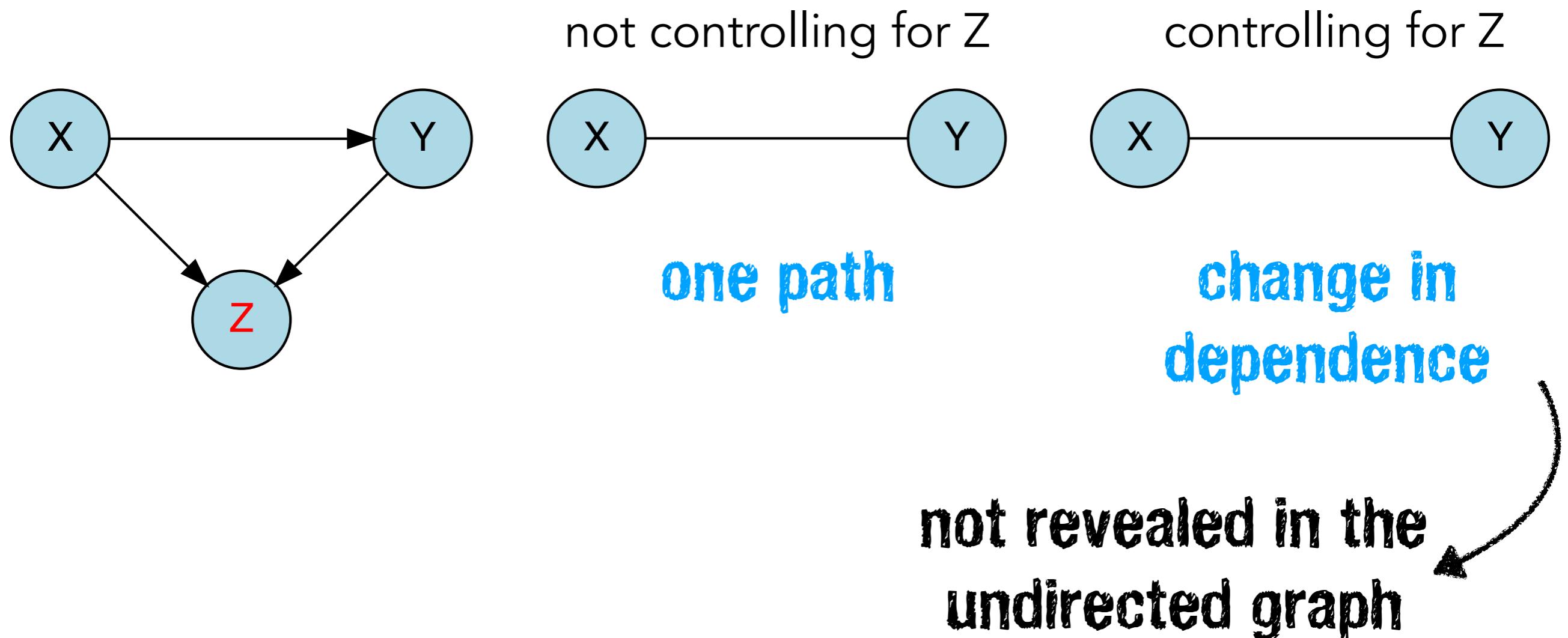
Z is a **bad** control here!

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



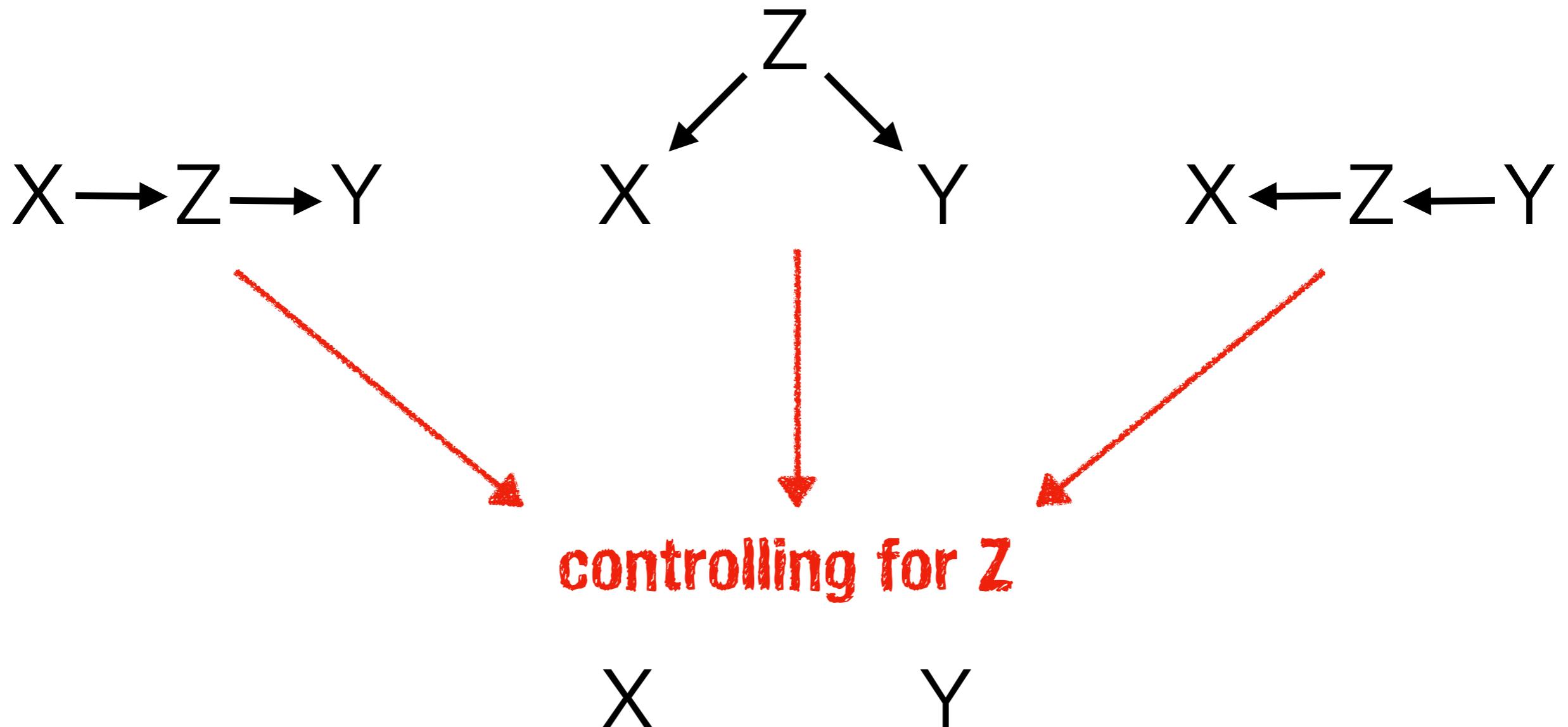
Z is a **bad** control here!

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 ...

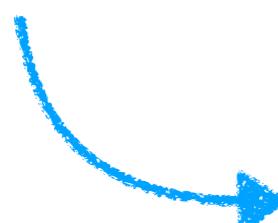


we need to manipulate X experimentally to tell these apart*

* sort of (see next slide)

When should I control for variables?

- causal discovery is a very active field



**what causal claims can we make
from observational data?**

Identifiability of Gaussian structural equation models with equal error variances

Jonas Peters*

Seminar for Statistics
ETH Zurich
Switzerland

Peter Bühlmann*

Seminar for Statistics
ETH Zurich
Switzerland

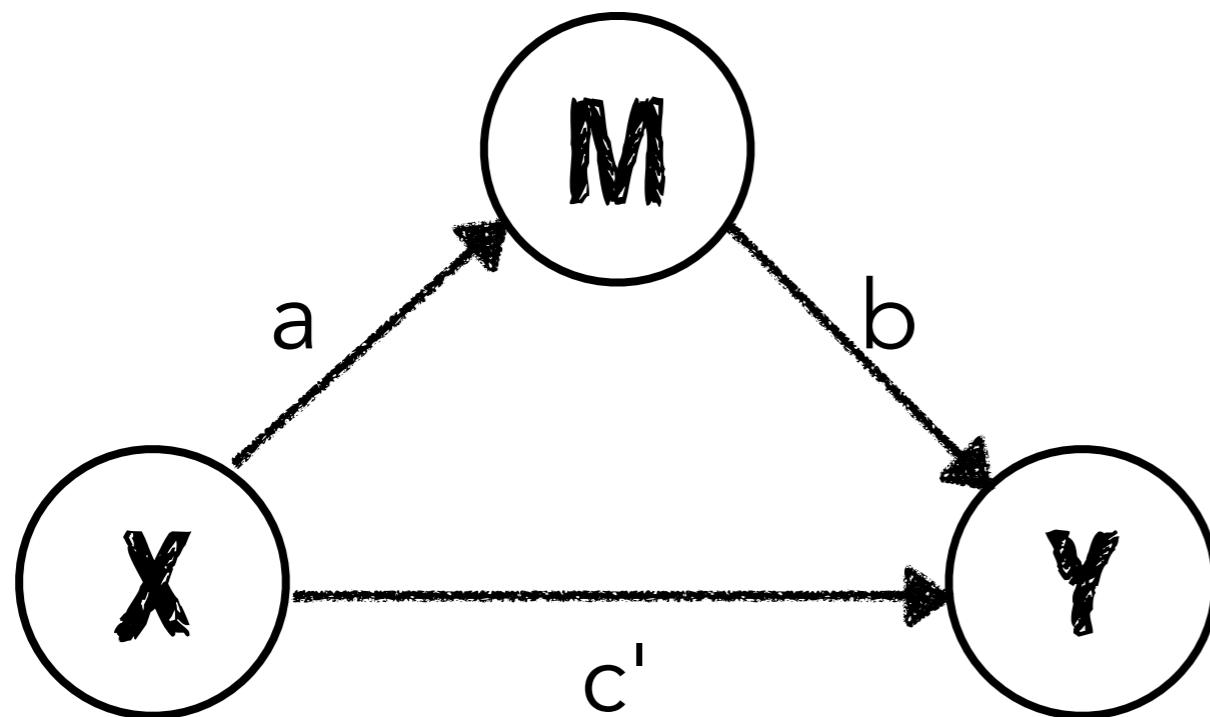
October 29, 2018

causal model is fully identifiable if all noise variables have the same variances, and all variables are observed

beyond the scope of our class ...

Mediation

Definition

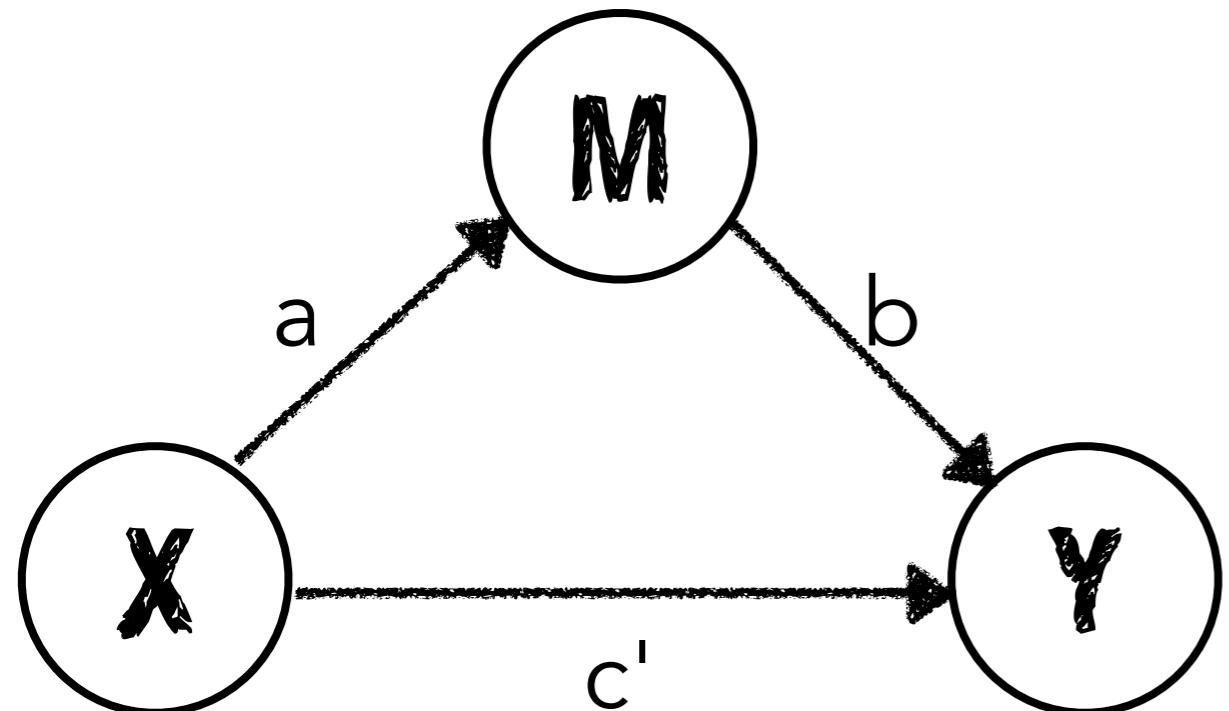


Rather than a direct causal relationship between **X** and **Y**, a mediation model proposes that **X** influences the mediator variable **M**, which in turn influences **Y**. Thus, the mediator variable serves to clarify the nature of the relationship between **X** and **Y**.

Adapted from Wikipedia

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

Example



X = grades in Psych 252

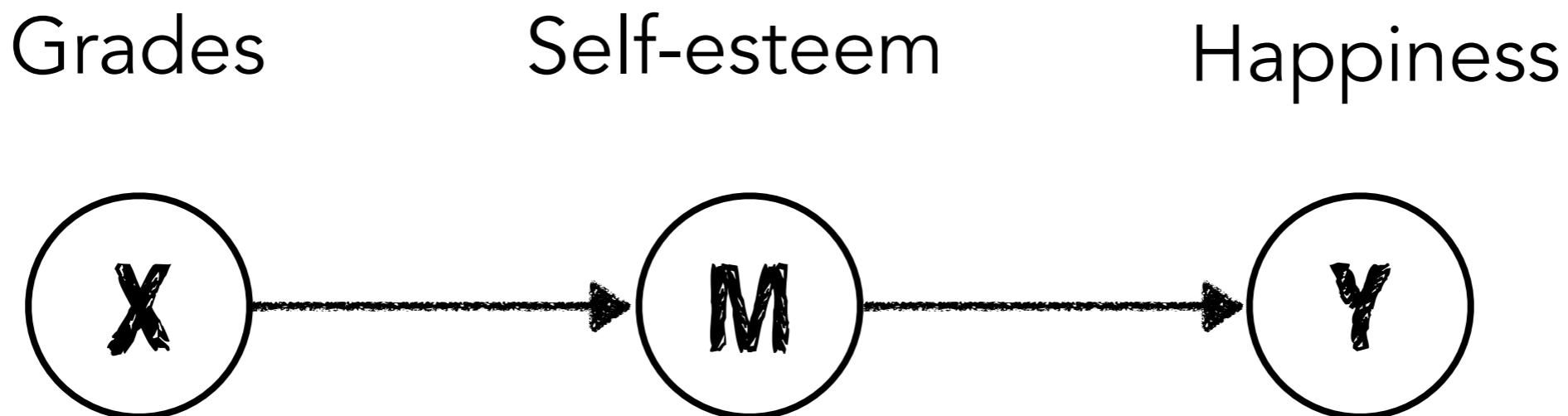
M = feelings of self-esteem

Y = happiness

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

Simulate a mediation analysis

```
1 # number of participants
2 n = 100
3
4 # generate data
5 df.mediation = tibble(
6   x = rnorm(n, 75, 7),           # grades
7   m = 0.7 * x + rnorm(n, 0, 5), # self-esteem
8   y = 0.4 * m + rnorm(n, 0, 5) # happiness
9 )
```



Bootstrapping

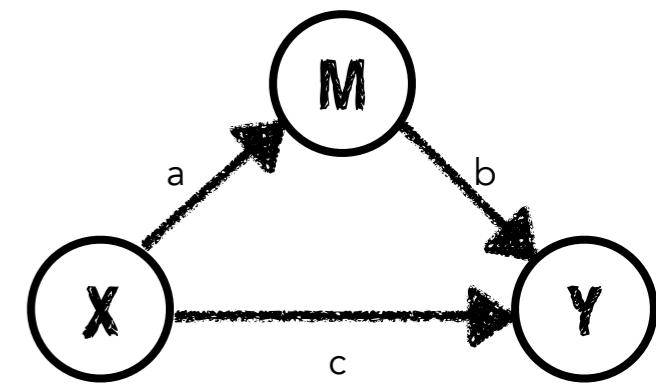
```
1 library("mediation")
```

```
2  
3 # bootstrapped mediation
```

```
4 fit.mediation = mediate(model.m = fit.m_x, ←  $\hat{m} = b_0 + b_1 \cdot x$ 
5 model.y = fit.y_mx, ←  $\hat{y} = b_0 + b_1 \cdot m + b_2 \cdot x$ 
6 treat = "x",
7 mediator = "m",
8 boot = T)
```

```
9  
10 # summarize results
```

```
11 fit.mediation %>% summary()
```



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

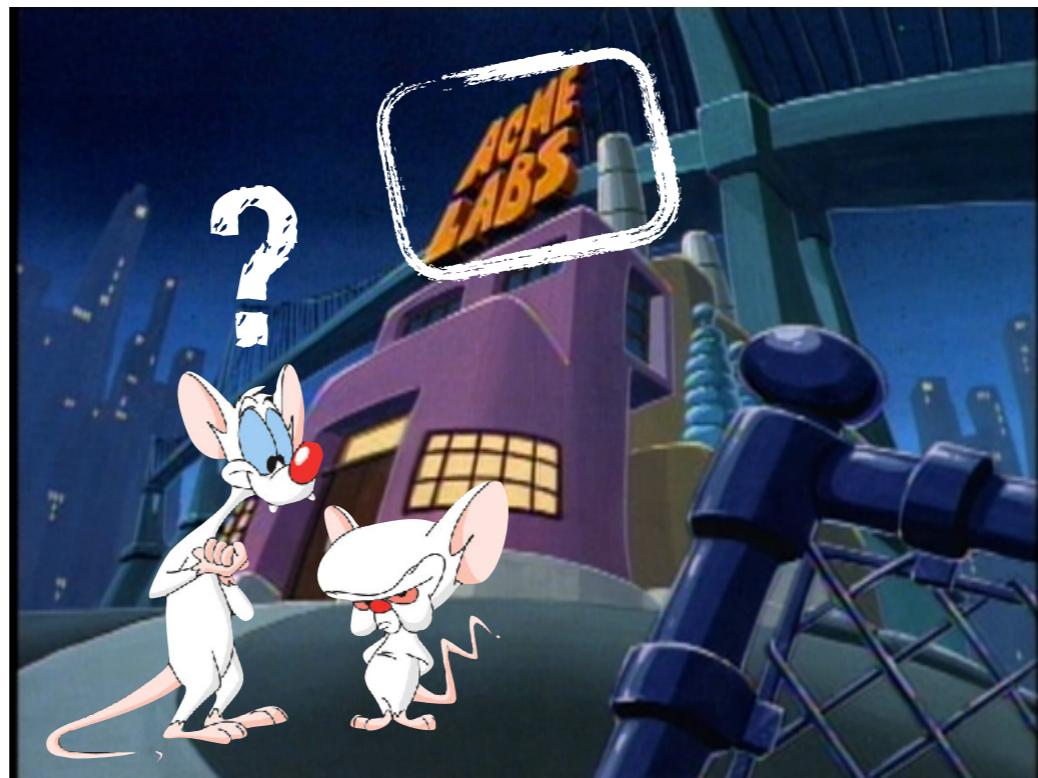
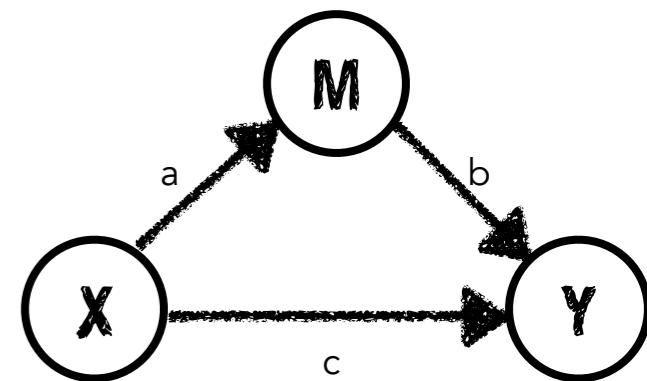
Nonparametric Bootstrap Confidence Intervals with the Percentile Method

	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

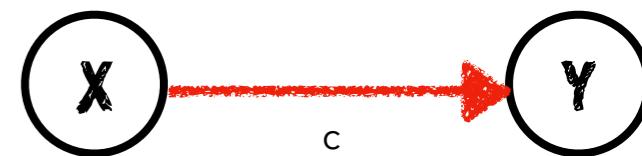
Sample Size Used: 100

Simulations: 1000



2. Bootstrapping

M



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)
```

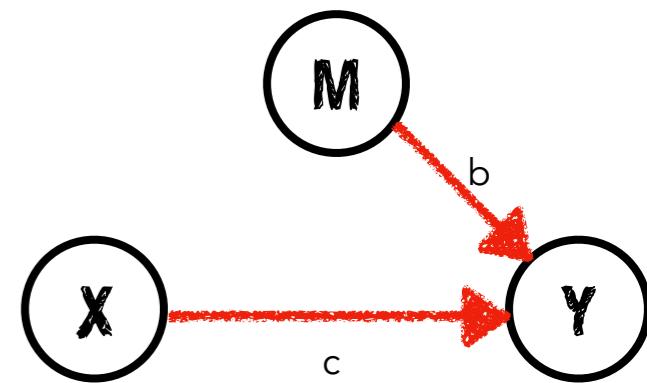
Residuals:

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 ' '
Sample Size Used:	100				

Simulations: 1000

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

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 ' '
Sample Size Used:	100				

Simulations: 1000

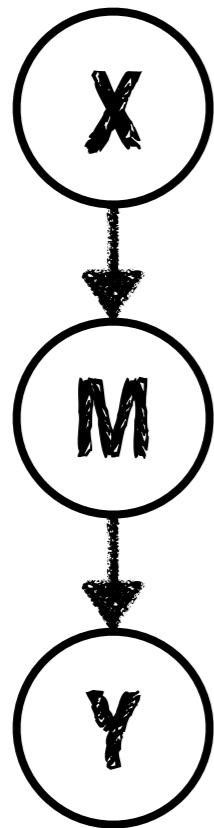
ACME: Average causal mediation effect

ACME = Total effect - ADE

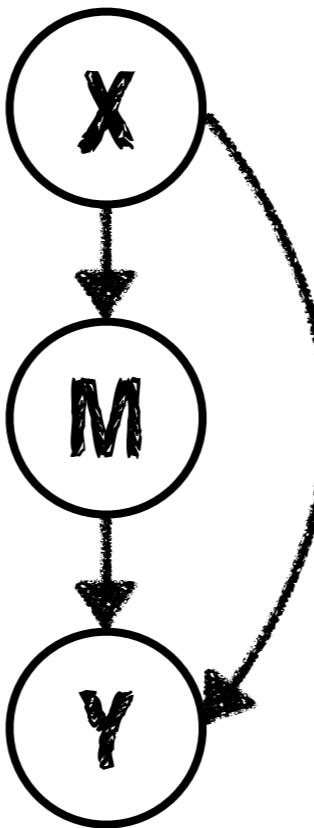
ADE: Average direct effect

Underlying causal model

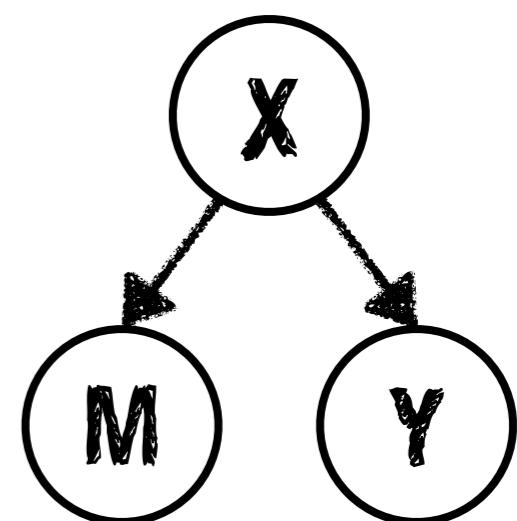
Full mediation



Partial mediation



No mediation

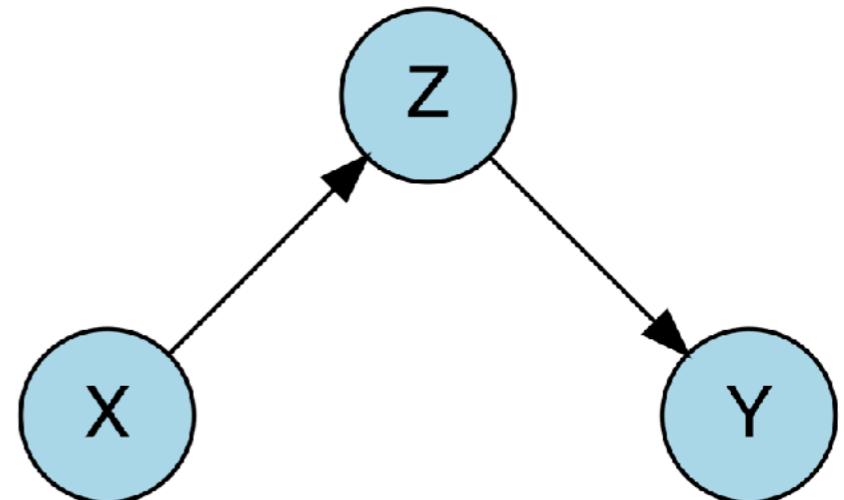


Full mediation: When the effect of **X** on **Y** completely disappears, **M** fully mediates between **X** and **Y**.

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!!!

```
1 set.seed(1)
2
3 n = 100 # number of observations
4
5 # causal chain
6 df.causal_chain = tibble(x = rnorm(n, 0, 1),
7                           z = 2 * x + rnorm(n, 0, 1),
8                           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

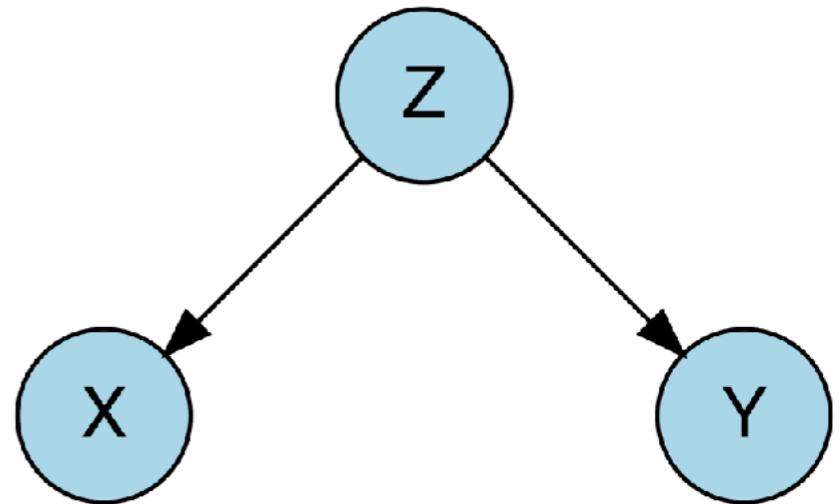
Sample Size Used: 100

Simulations: 1000

nice mediation result!

Beware of mediation analyses!!!

```
1 set.seed(1)
2
3 n = 100 # number of observations
4
5 # common cause
6 df.common_cause = tibble(z = rnorm(n, 0, 1),
7                           x = 2 * z + rnorm(n, 0, 1),
8                           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

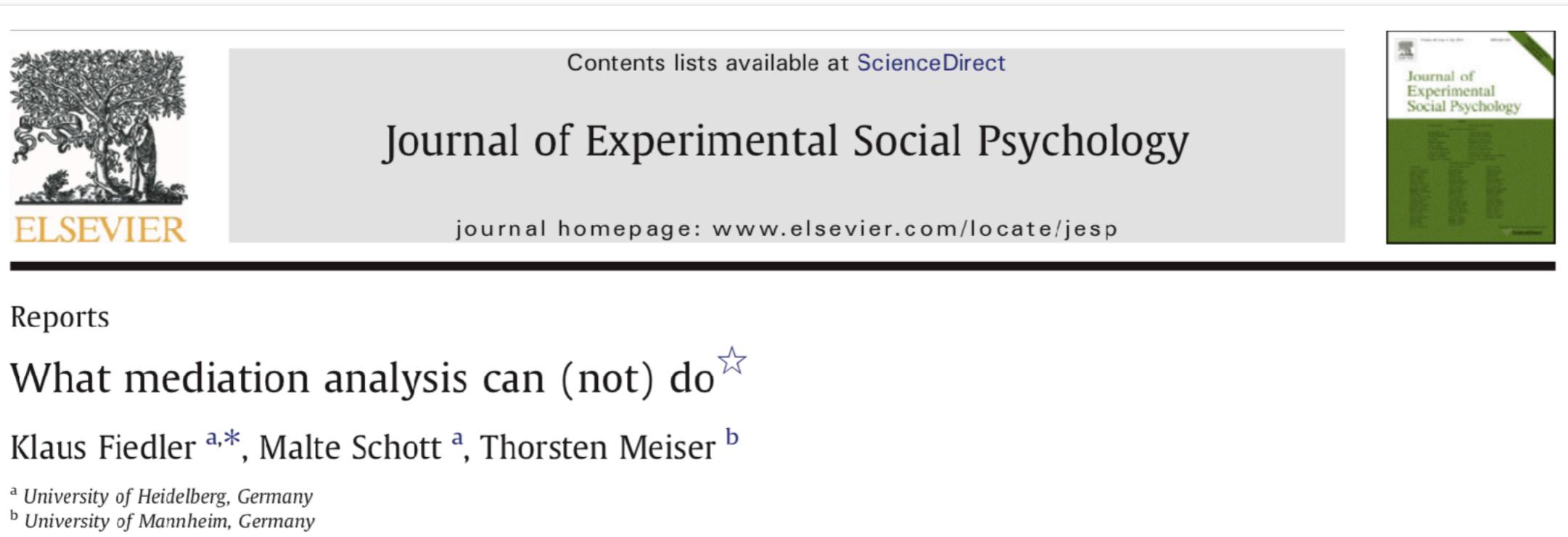
Sample Size Used: 100

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



The image shows the cover of a journal article from the Journal of Experimental Social Psychology. The Elsevier logo is on the left, featuring a tree and the word 'ELSEVIER'. The title 'Journal of Experimental Social Psychology' is in the center, with 'Contents lists available at ScienceDirect' above it and 'journal homepage: www.elsevier.com/locate/jesp' below it. To the right is a thumbnail of the journal's cover page.

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. 111

Limitations

many-to-one mapping

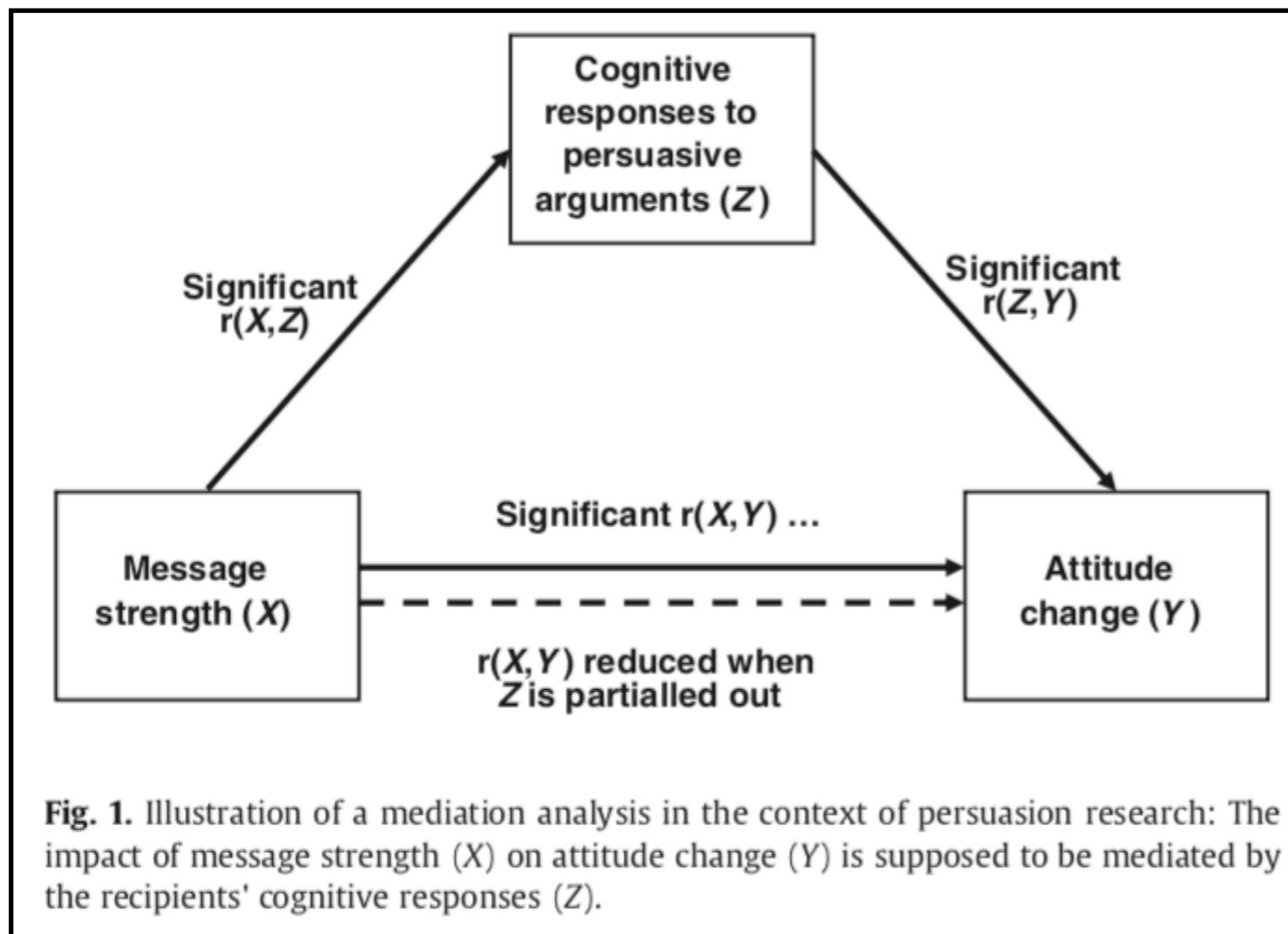


Fig. 1. Illustration of a mediation analysis in the context of persuasion research: The impact of message strength (X) on attitude change (Y) is supposed to be mediated by the recipients' cognitive responses (Z).

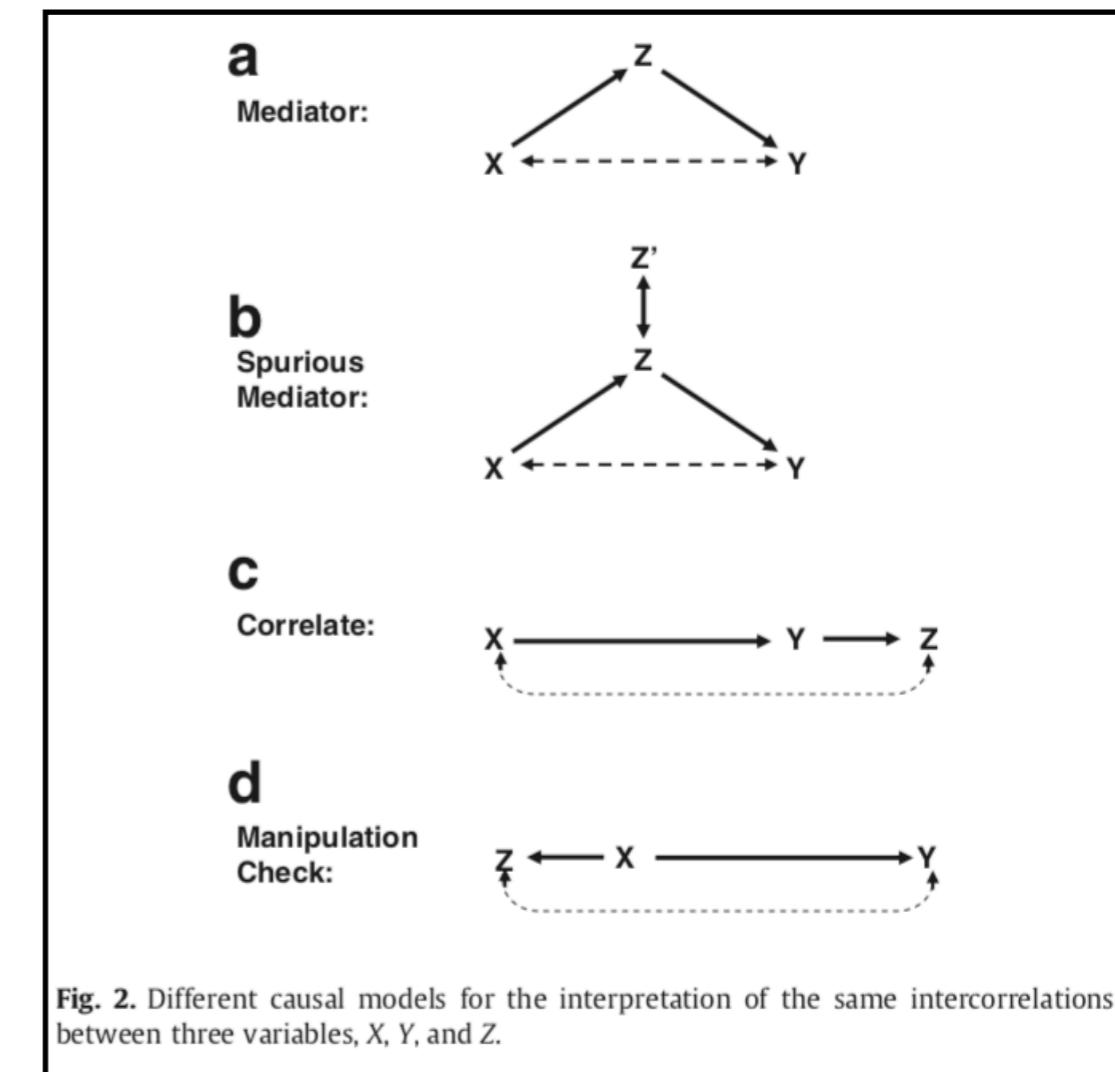


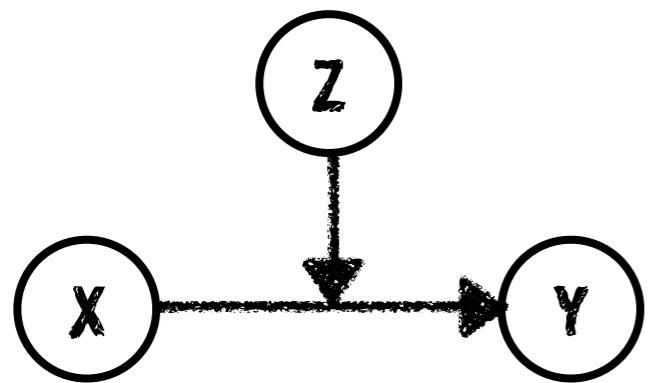
Fig. 2. Different causal models for the interpretation of the same intercorrelations between three variables, X , Y , and Z .

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. 112

Moderation

Definition

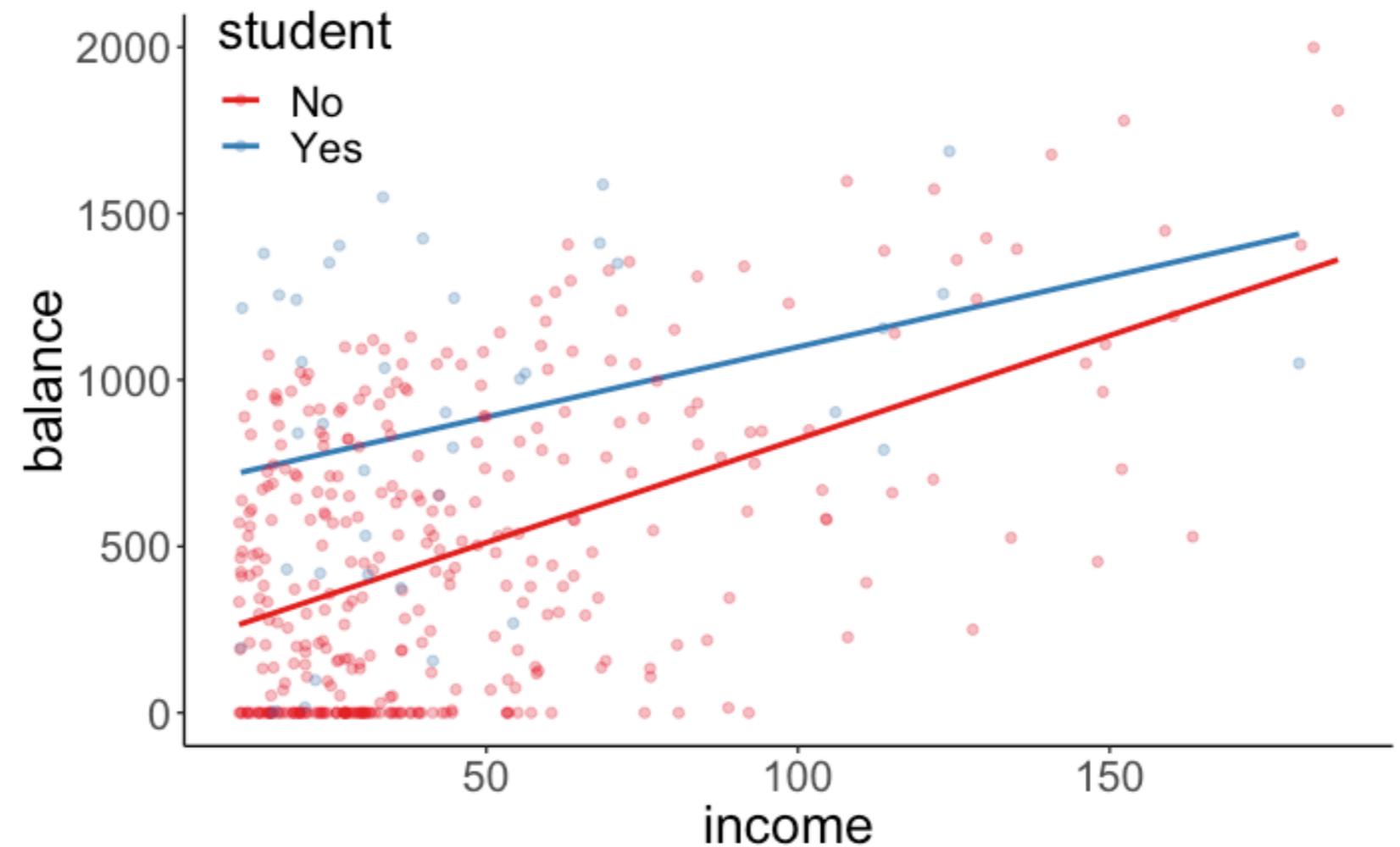


Moderation means that the effect of a predictor depends on the value of another.

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.



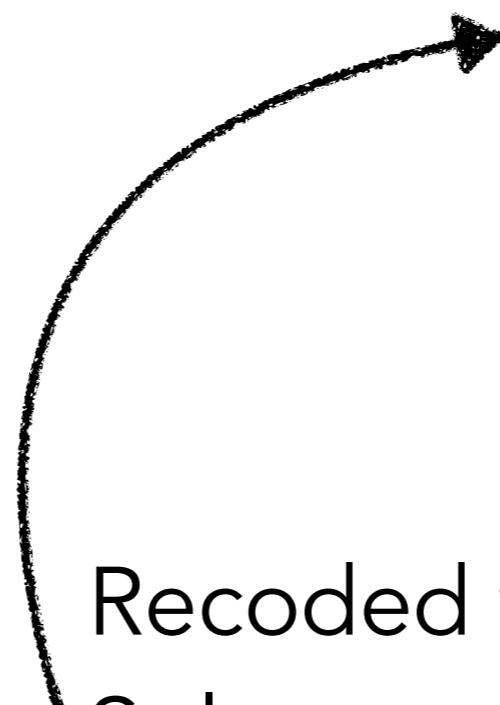
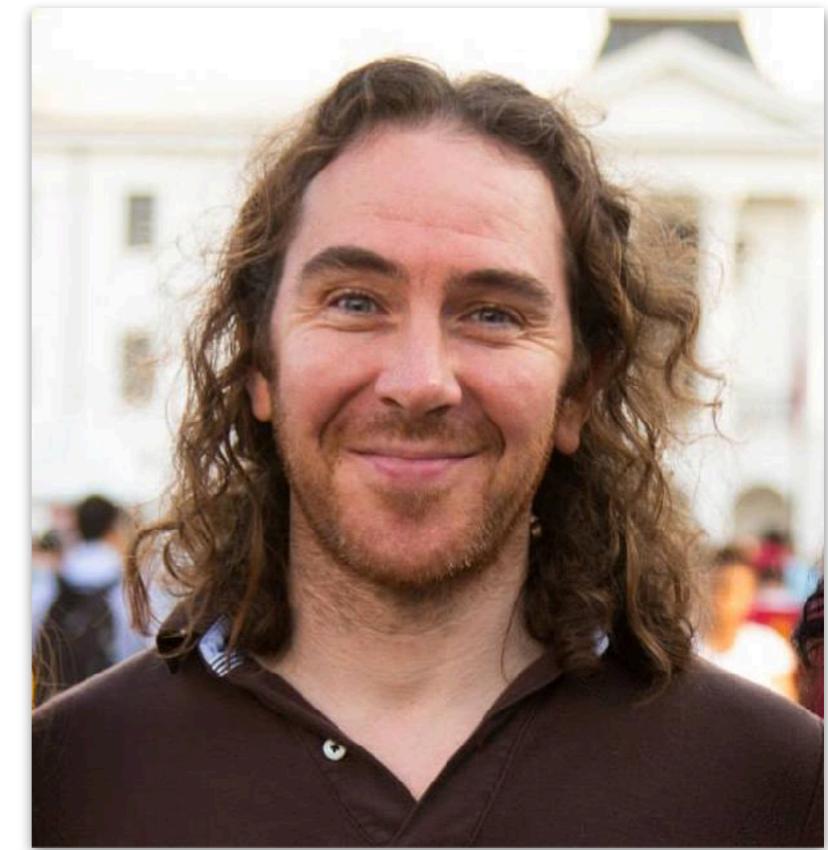
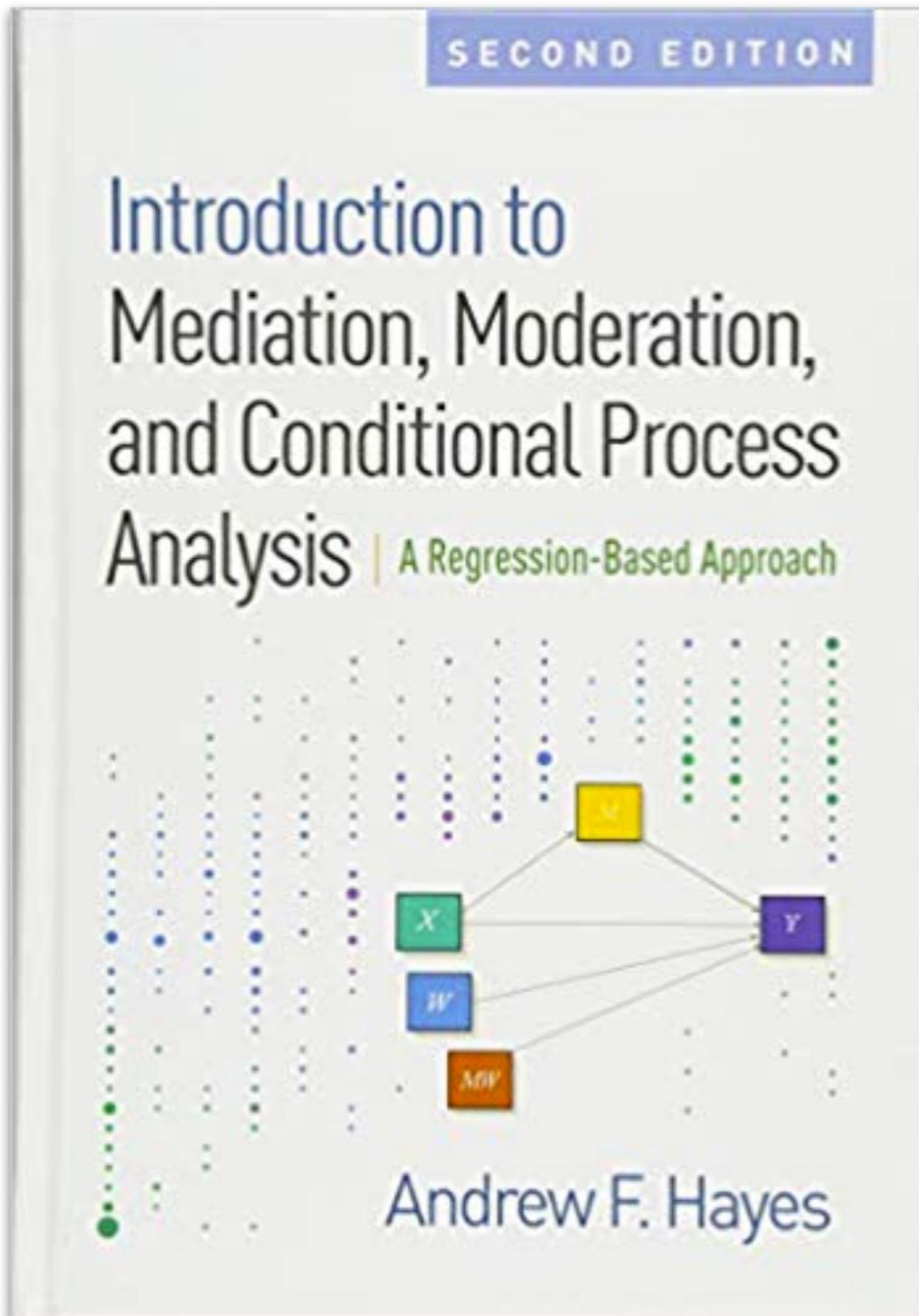
$$\widehat{\text{balance}}_i = 200.62 + 6.22 \cdot \text{income}_i + 476.68 \cdot \text{student}_i - 2.00 \cdot (\text{income}_i \times \text{student}_i)$$

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

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](https://bookdown.org/connect/#/apps/1523/access)

Summary

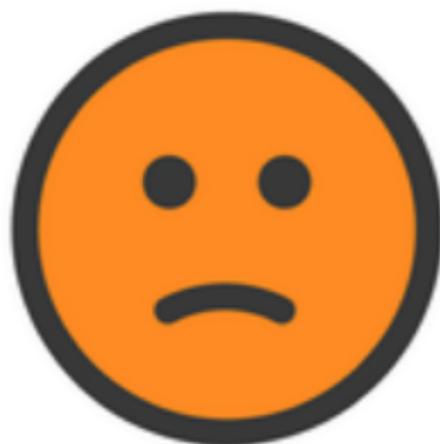
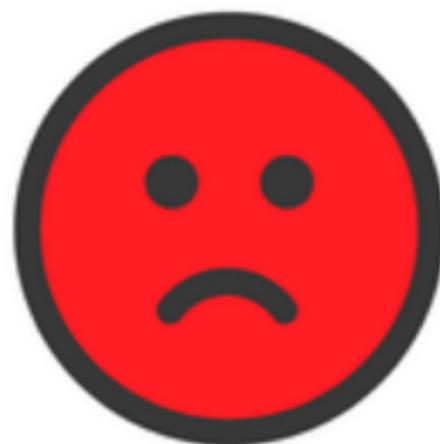
- Quick recap
- Model comparison
 - Cross-validation in action
 - AIC and BIC
- Causation vs. correlation
- Controlling for variables
- Mediation
- Moderation

Feedback

How was the pace of today's class?

much a little just a little much
too too right too too
slow slow fast fast

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!