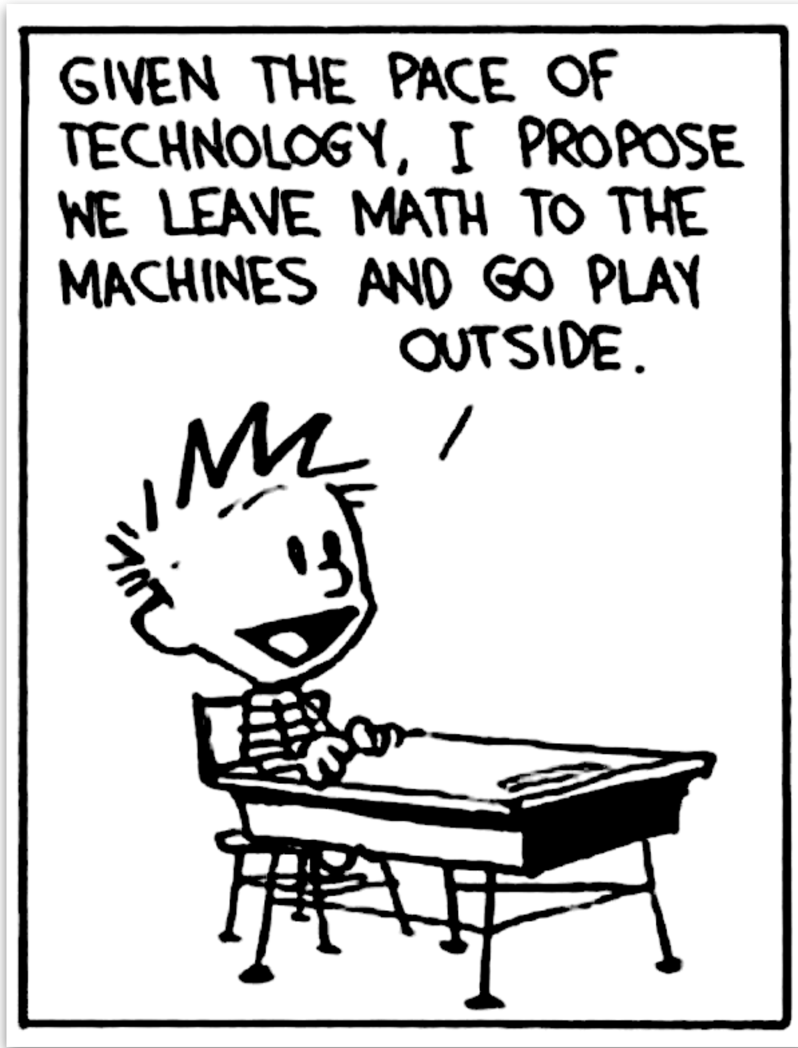
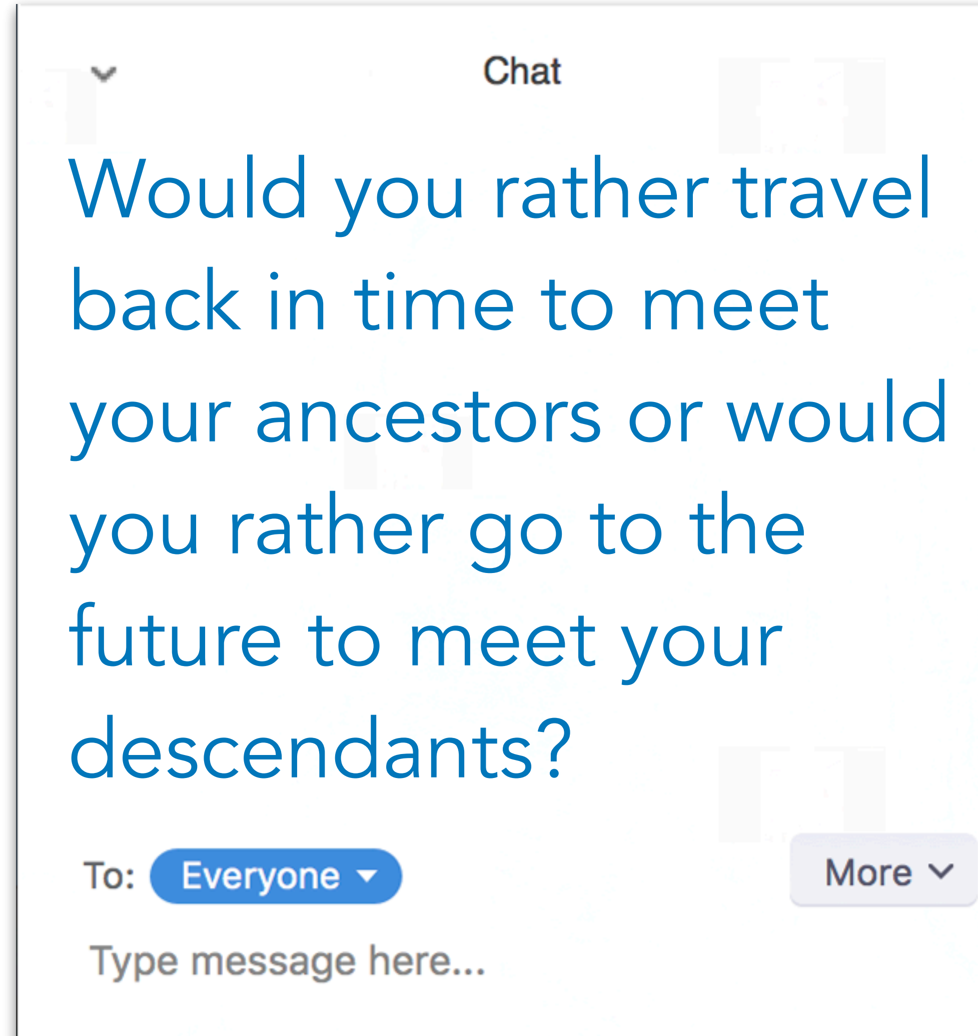
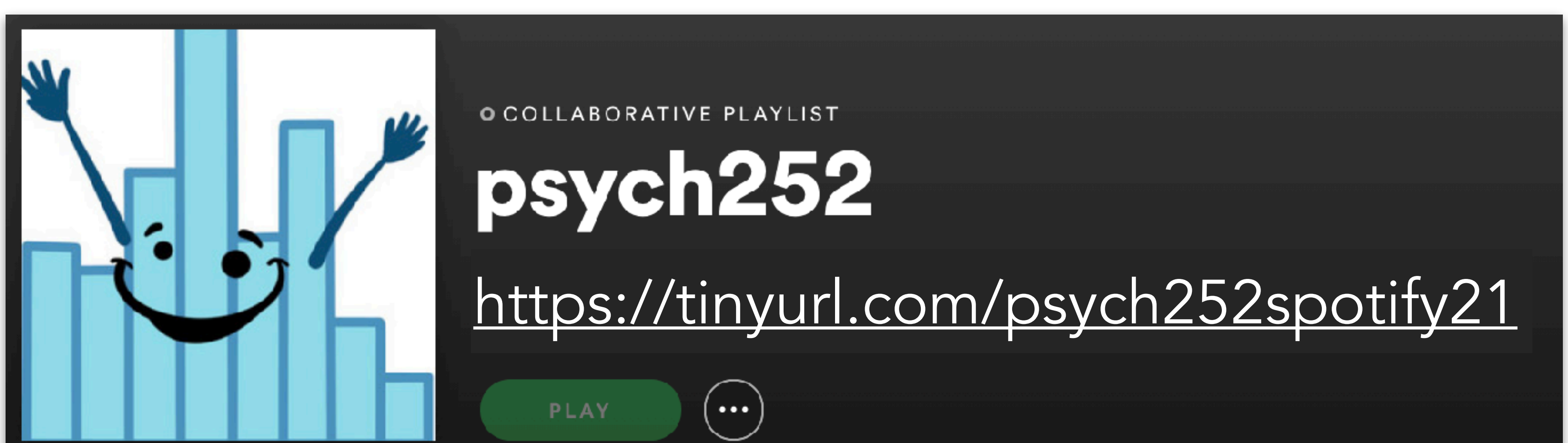
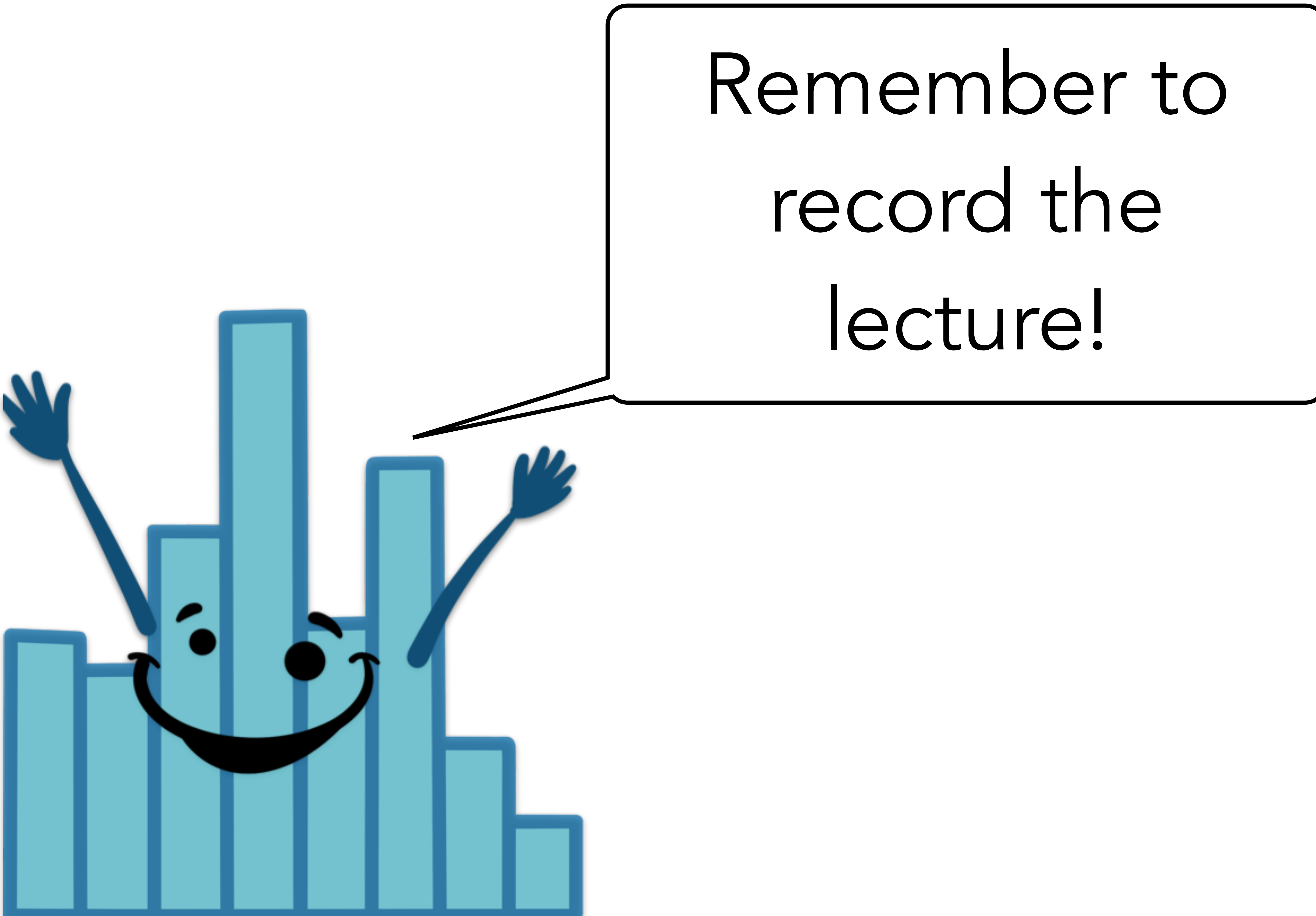
Mediation and Moderation



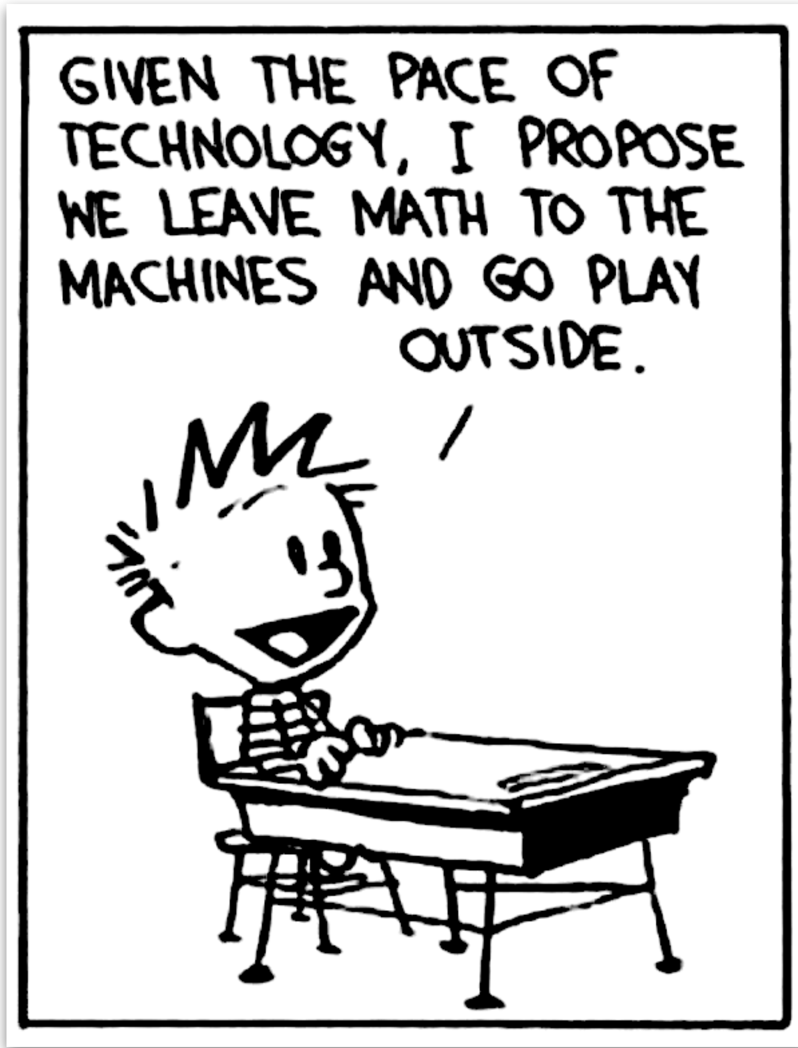


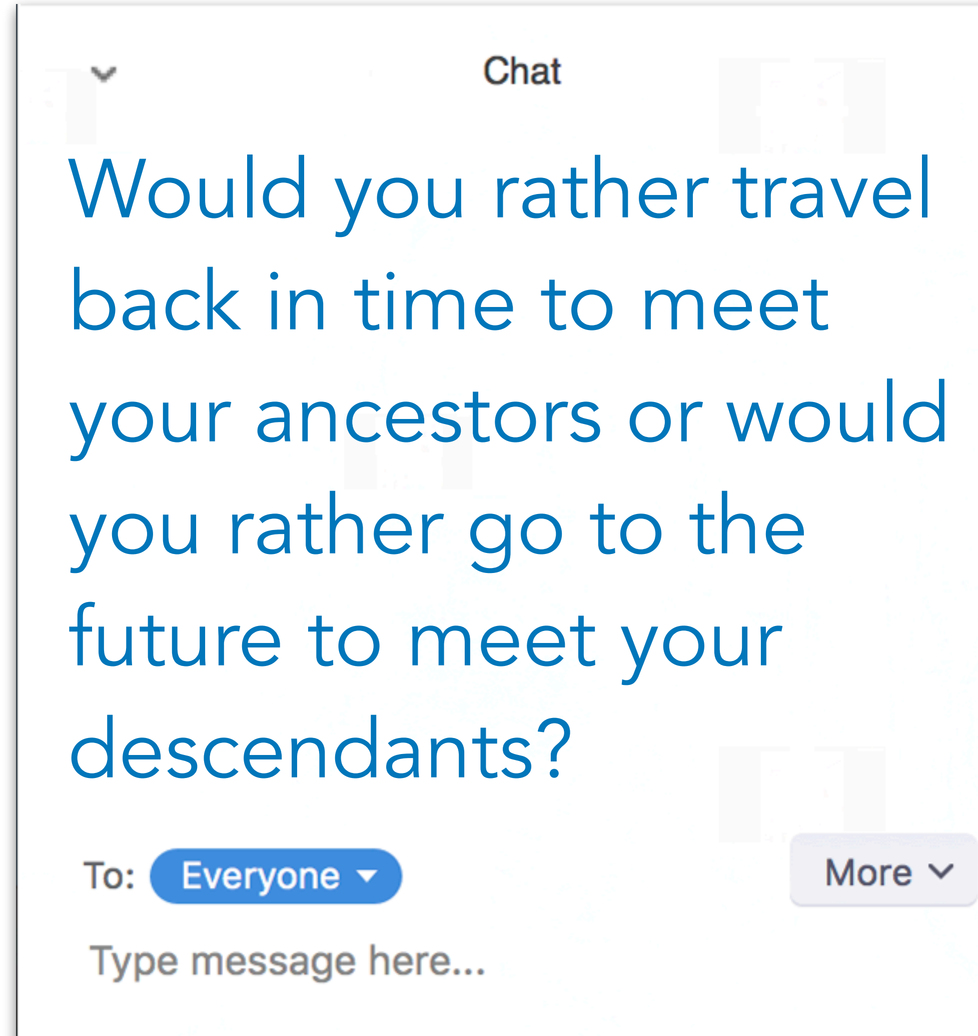


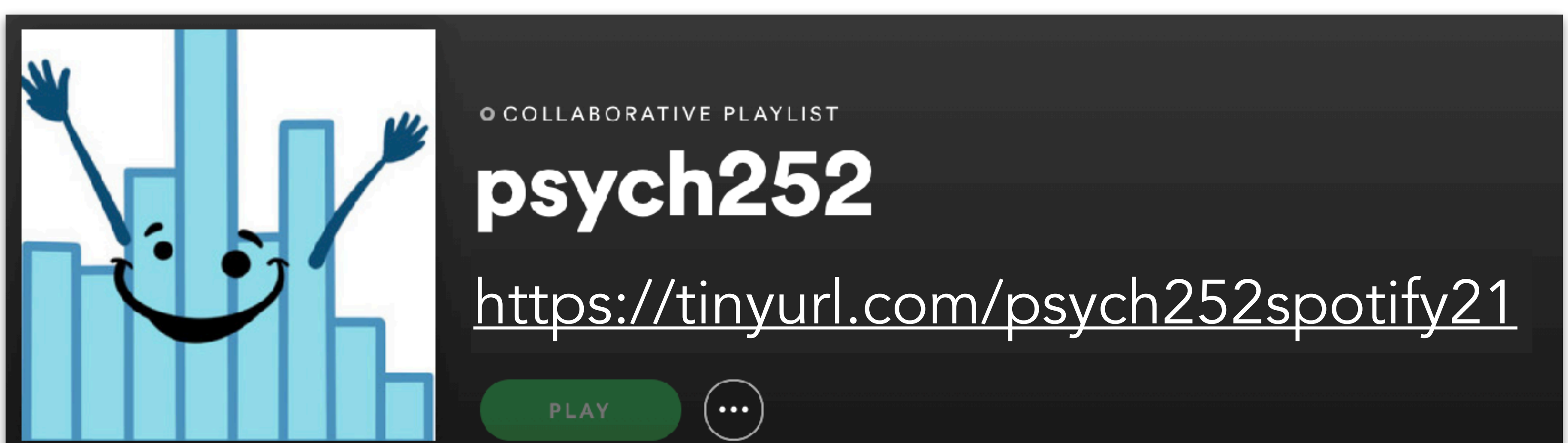
02/19/2021


Remember to record the lecture!

Mediation and Moderation







02/19/2021

Things that came up

gganimate

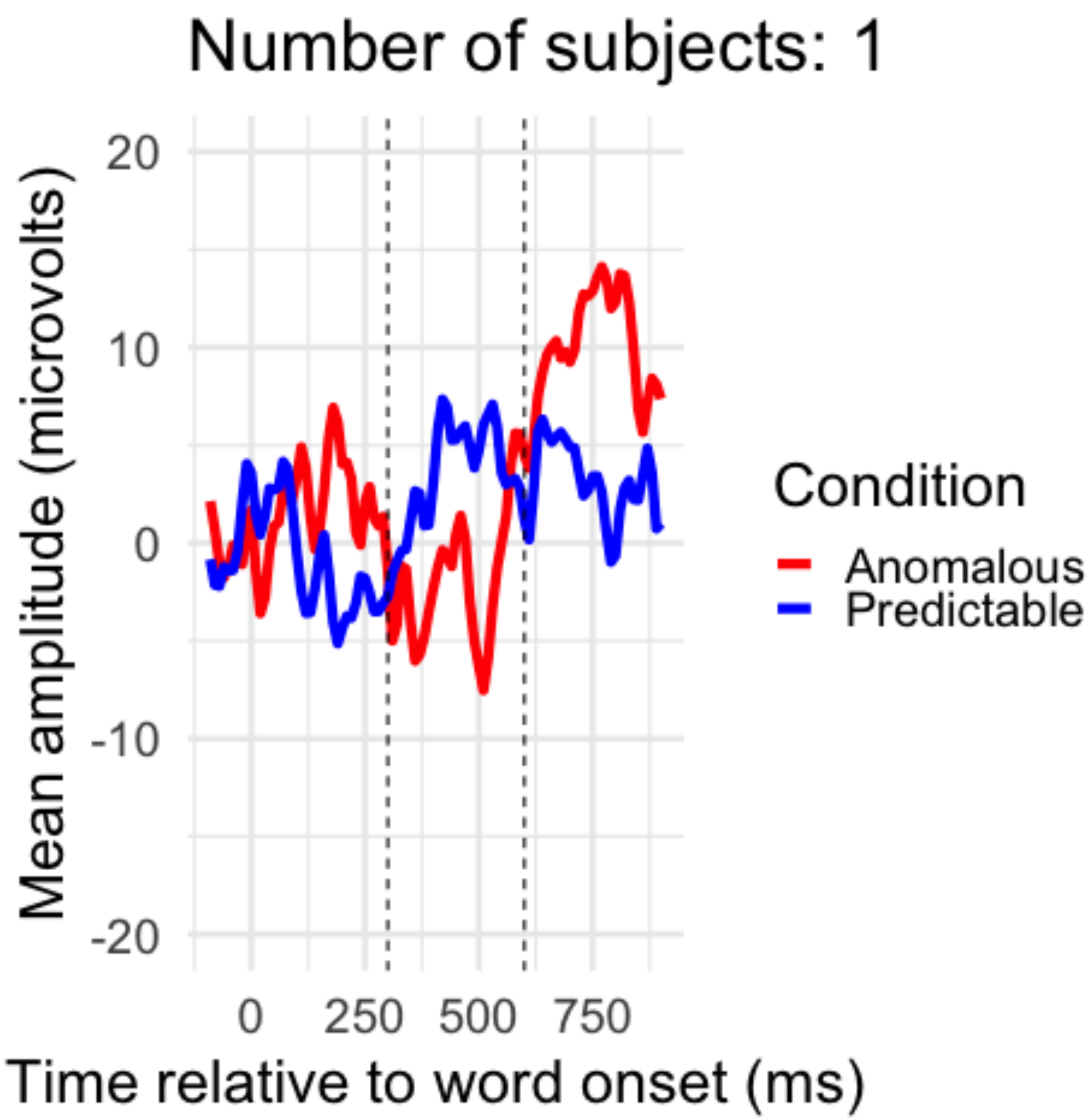
Great use of `gganimate`

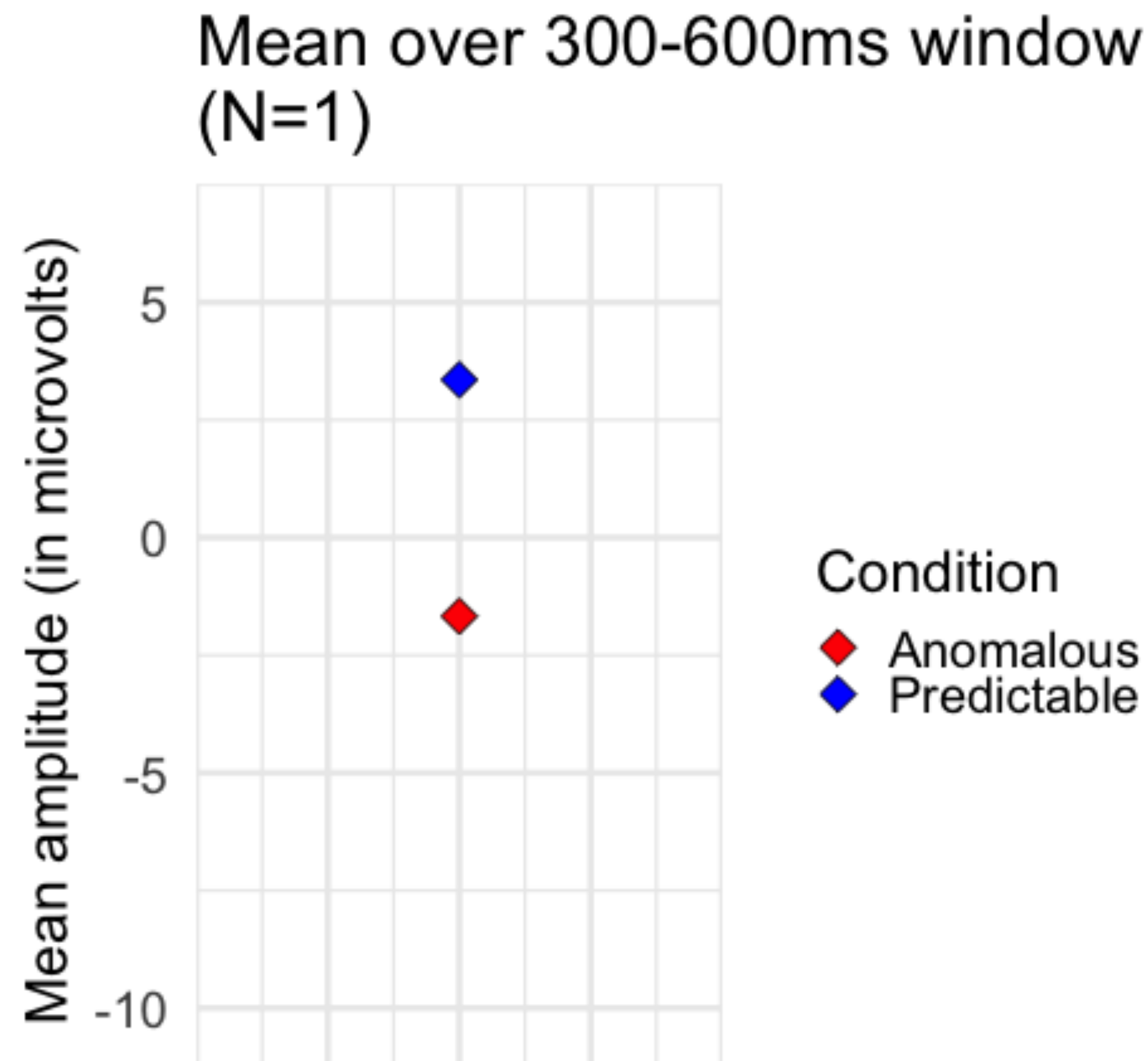
Rachel A. Ryskin

`ryskin@mit.edu`

Curriculum Vitae

Google Scholar | GitHub | Twitter

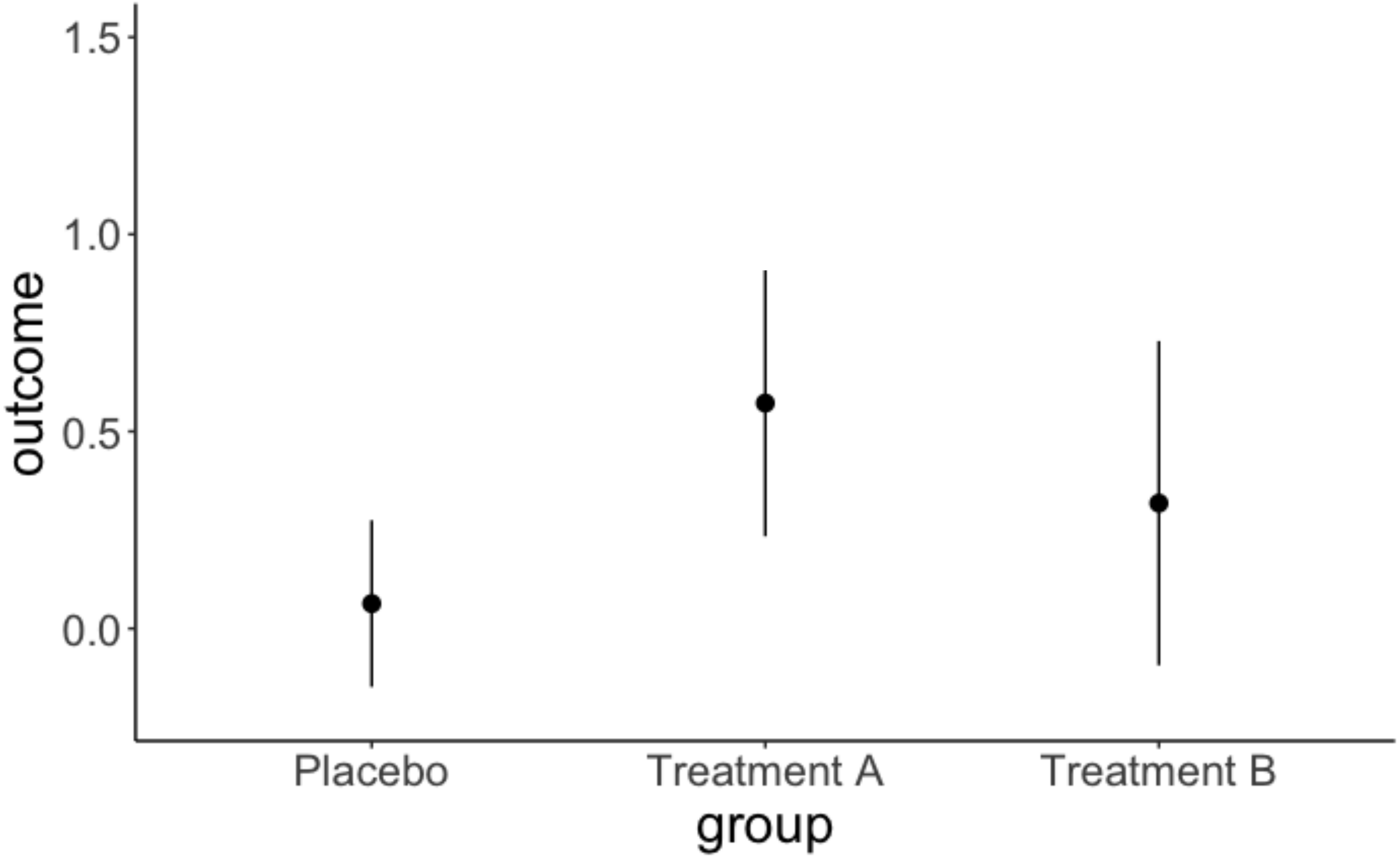




difference in significance vs. significant differences

significant differences

"Our tendency to look for a difference in significance should be replaced by a check for the significance of the difference."

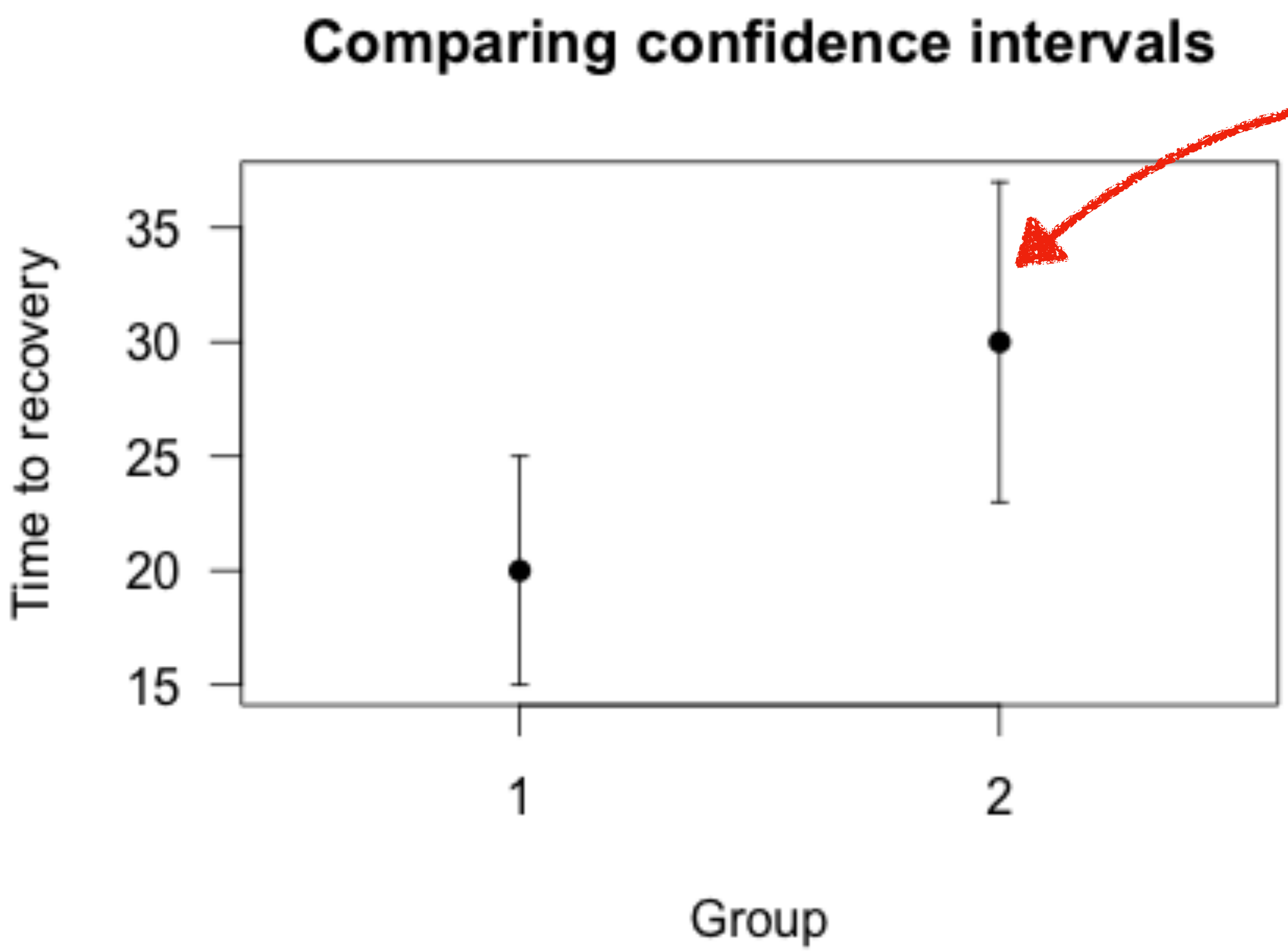


"We compared treatments A and B with a placebo. Treatment A showed a significant benefit over placebo, while treatment B had no statistically significant benefit. Therefore, treatment A is better than treatment B."

https://www.statisticsdonewrong.com/significant-differences.html

significant differences

Significant difference between Group 1 and 2?



what do the error bars mean?

standard deviation?

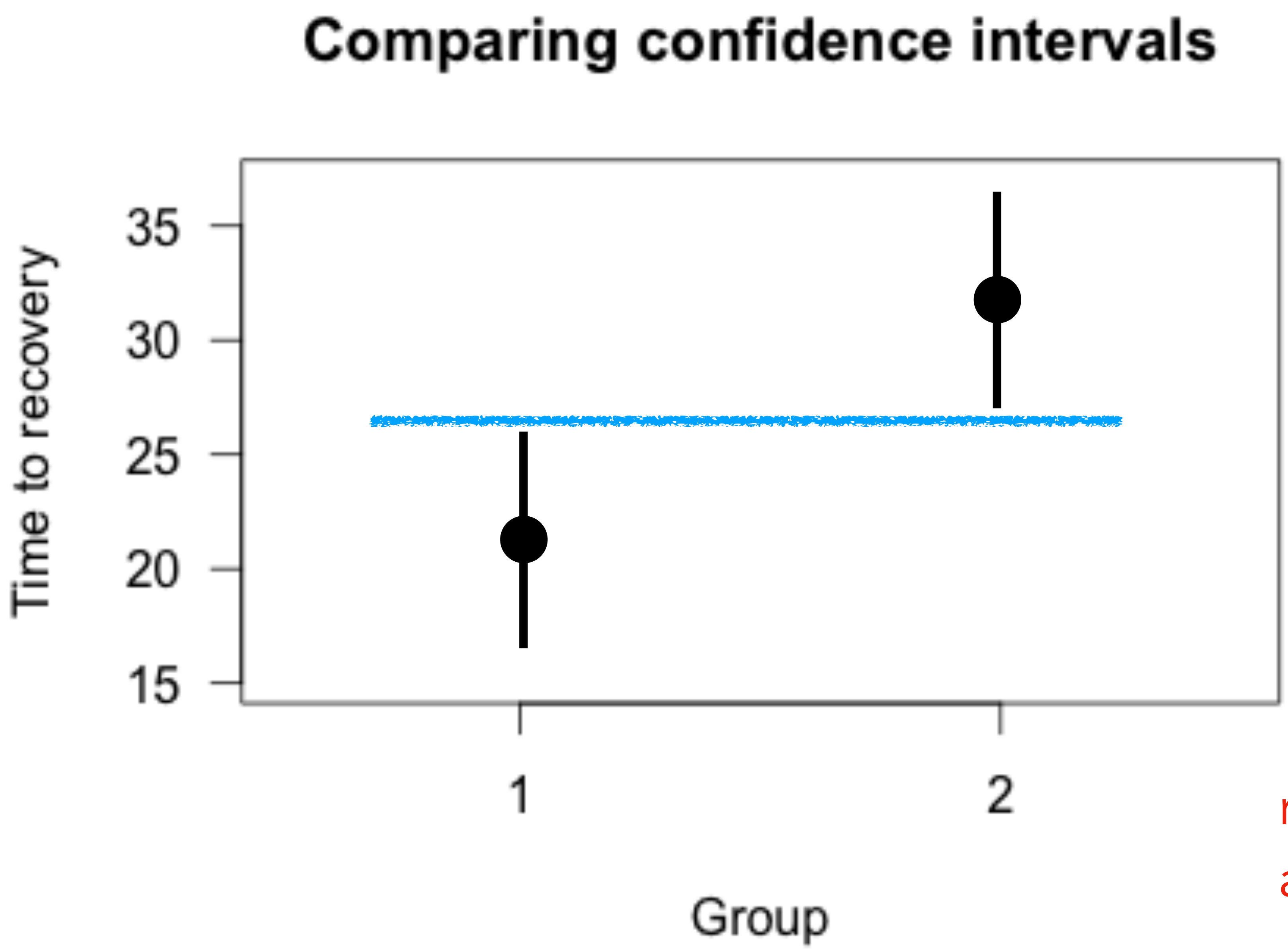
standard error of the mean?

confidence intervals?

What we would like to know, does the confidence interval of the difference between groups exclude 0?

significant differences

Significant difference between Group 1 and 2?



95% confidence intervals don't overlap

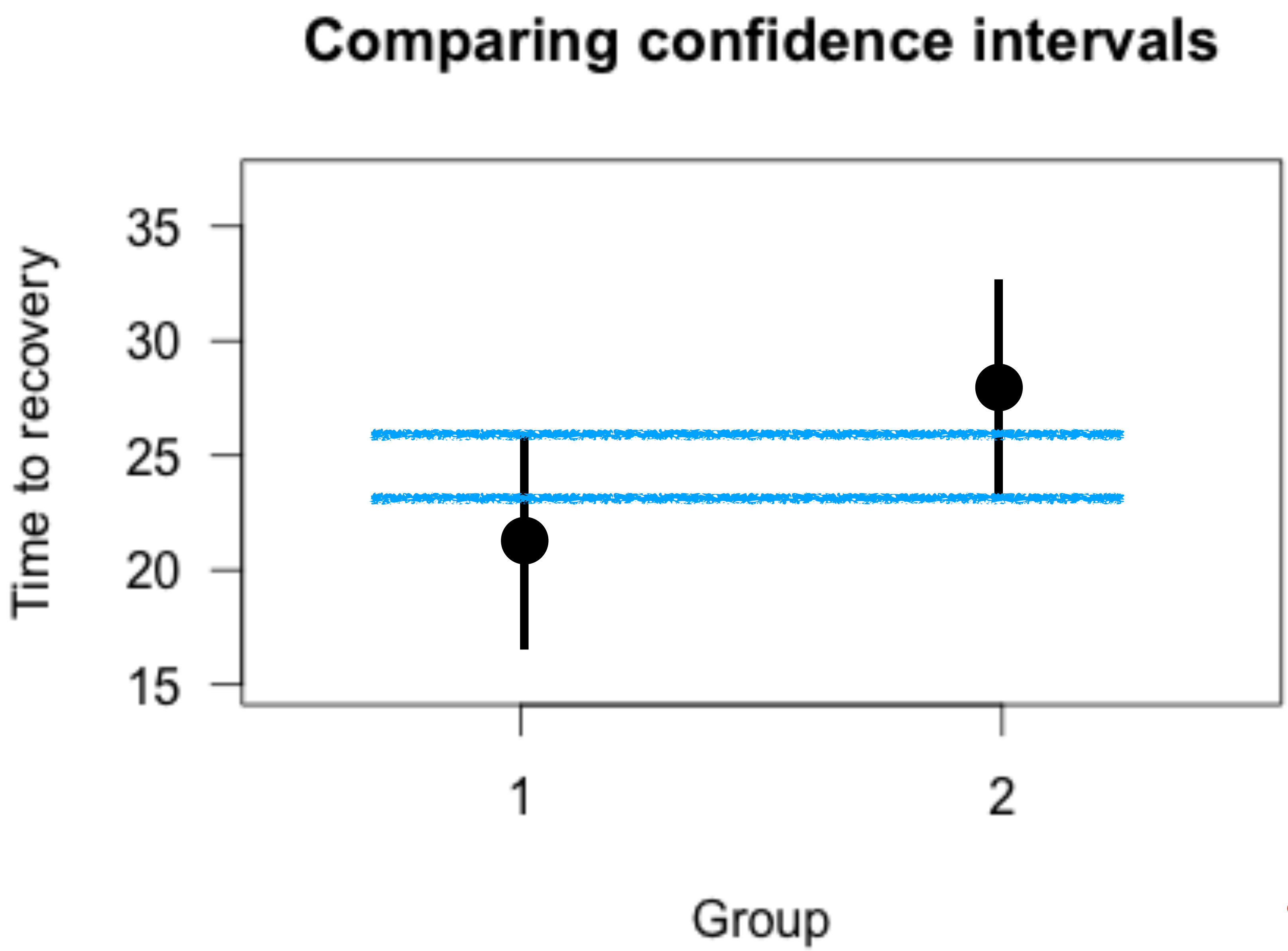
conservative test of significance

rejects H_0 **less often** than the appropriate statistical procedure

Schenker & Gentleman (2001) On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*

significant differences

Significant difference between Group 1 and 2?



95% confidence intervals don't overlap with mean

anti-conservative test of significance

rejects H_0 **more often** than the appropriate statistical procedure

Schenker & Gentleman (2001) On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*

Logistics

Homework 4

Homework 4

Graded and solutions are on Canvas.

When poll is active, respond at **PollEv.com/psych252**

How many hours did it take you to complete Homework 4?

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15 or more

Total Results: 0

Midterm

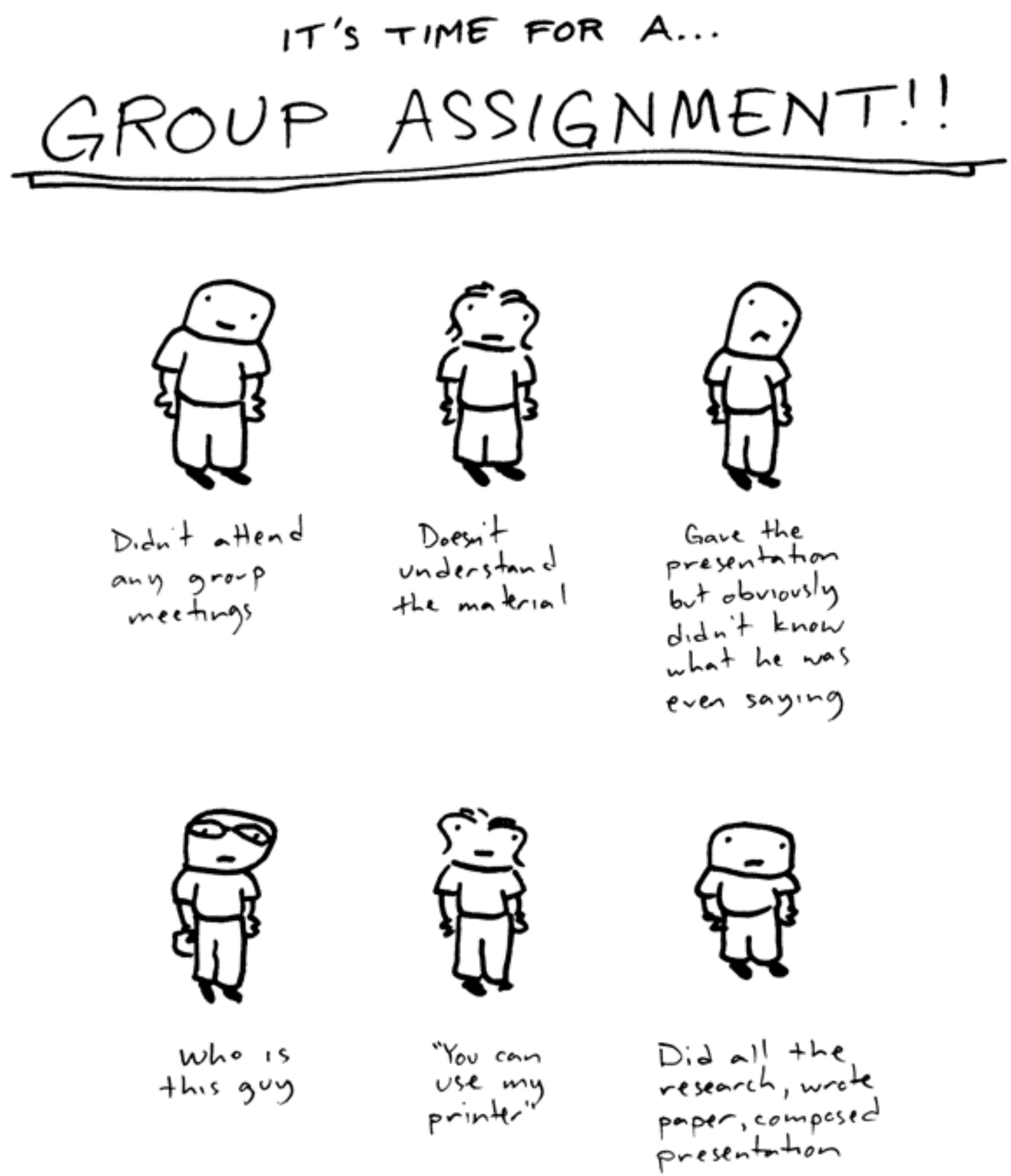
How many hours did it take you to complete the midterm? (just write one number: e.g. 9 if it took you 9 hours)

Project proposal

Project proposal

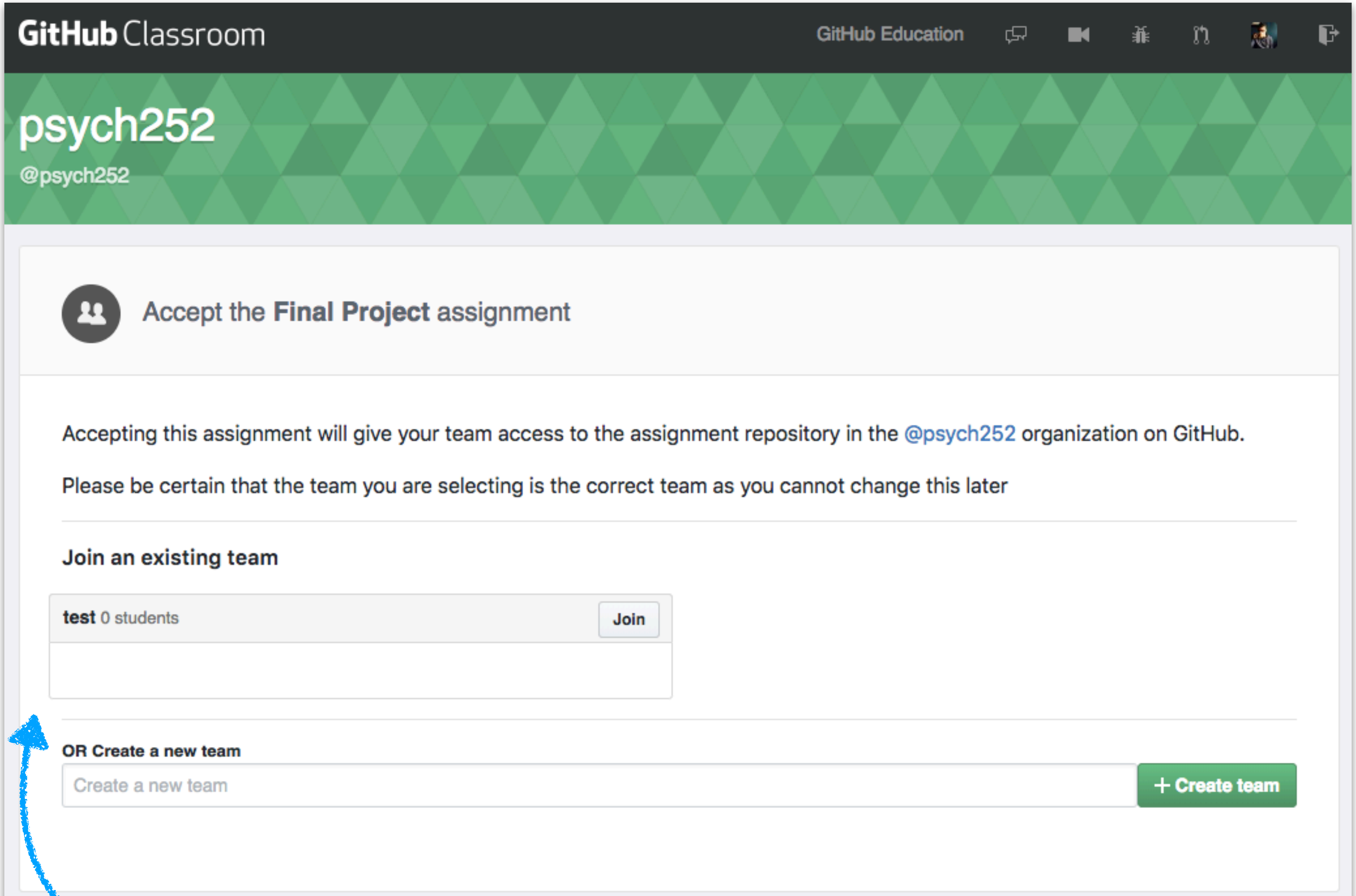
Members of teams will all the get same grade!

maximum 3 team members



Toothpaste For Dinner.com

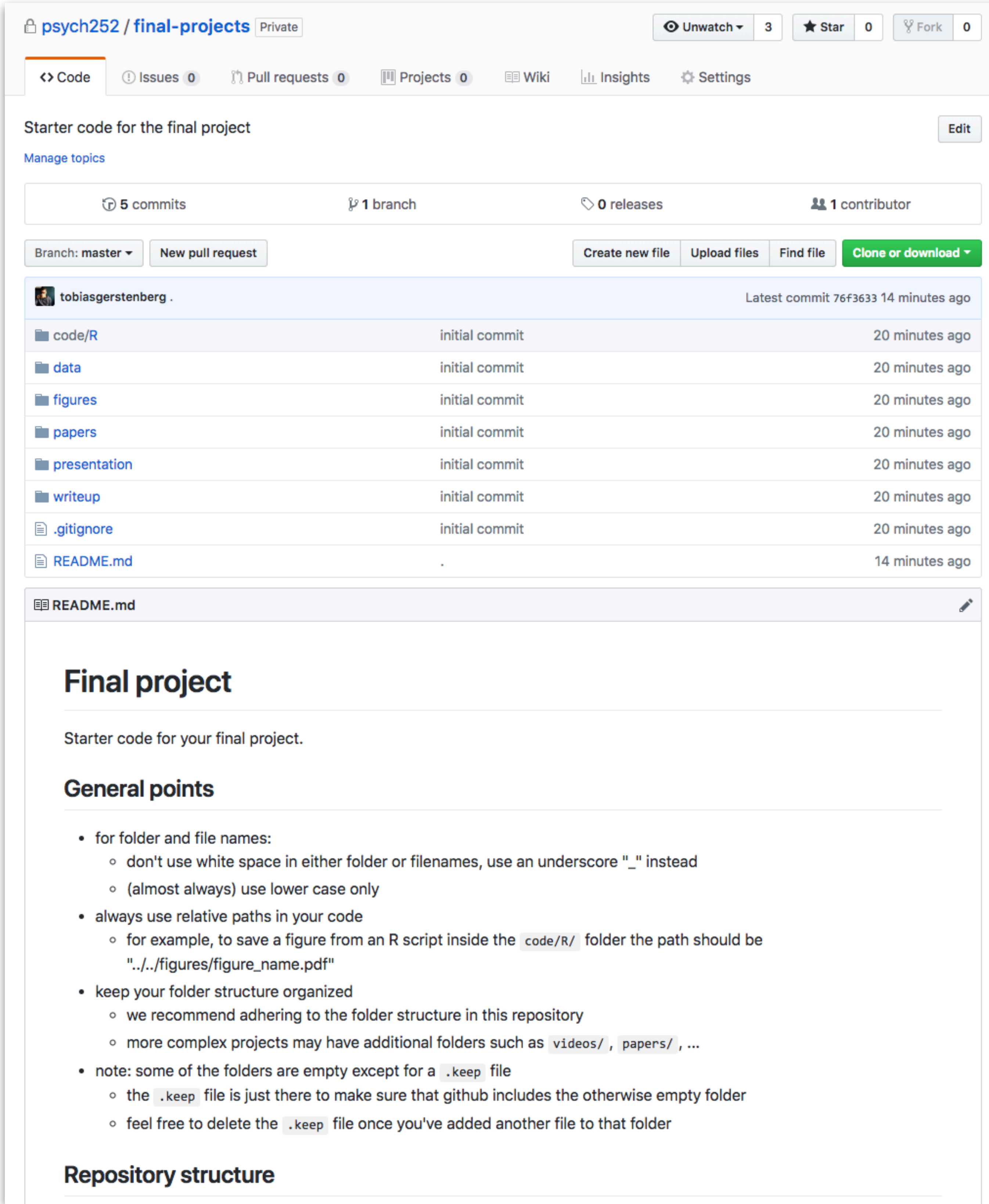
Project proposal



join a team or make a new one

https://classroom.github.com/g/A9487wbl

Project proposal



- 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 Piazza in case you experience any problems getting set up

Project proposal

RMarkdown template

project_proposal.Rmd

Knit Run

```
---
title: "A catchy project title goes here"
subtitle: "My team's name goes here"
author: "The team members' names go here"
date: "`r Sys.time()`"
urlcolor: blue # to show hyperlinks in blue when printed as pdf

# edit the output format below
# output: html_document # use this to render to html
output: pdf_document # use this to render to pdf

---

*Instructions*

The project proposal is due on __Thursday, February 25th at 8pm__. It should not be longer than __700 words__. It may contain code (code doesn't count toward the word limit).

# Research question (3 pt)

## What's your main research question? (1 pt)

> YOUR ANSWER HERE

## Which hypotheses are you trying to test? (2 pt)

_Enumerate 1-3 hypotheses in terms of directional relationships. For example, you might write something like "X is predicted to increase as Y increases", or "We predict that performance in Group 2 will be significantly better than performance in Group 1"._

> YOUR ANSWER HERE

# Methods (4 pt)
```

Upload the pdf to canvas

Project proposal

The project proposal is due on

Thursday, February 25th at 8pm

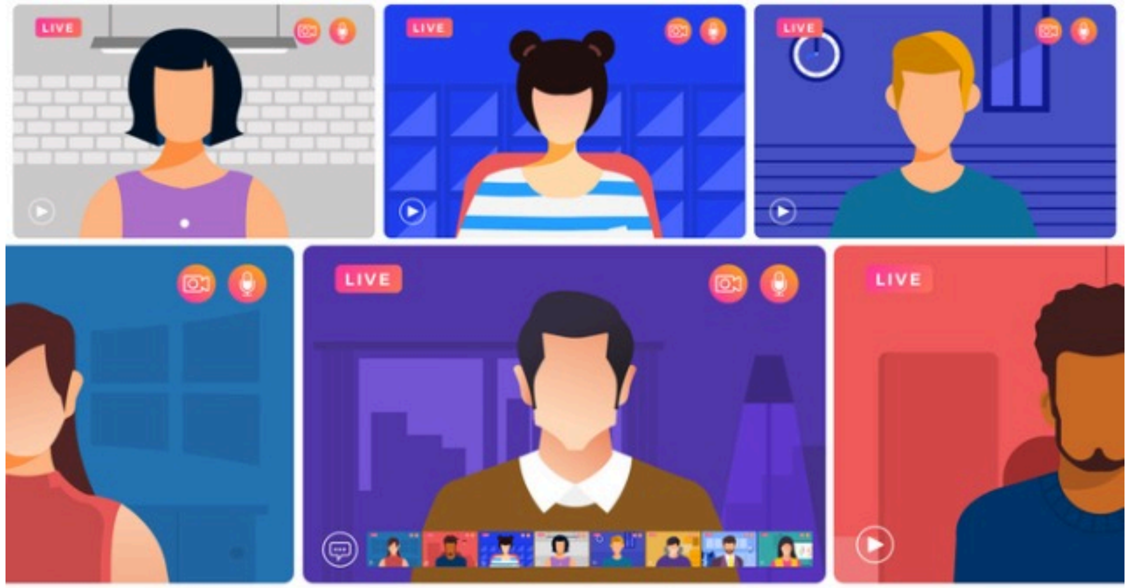
instructions of how to submit
will be released later today

Plan for today

- Controlling for variables
- Mediation
- Moderation

Controlling for variables

Breakout rooms



Tasks:

1. **What does "controlling" for variables mean?**

2. **When should I control for variables?**

Size: ~3 people

Time: 5 minutes

Report: Share your group's thinking in the main room.

Controlling for variables

1. What does "controlling" for variables mean?

2. When should I control for variables?

check this out

http://causality.cs.ucla.edu/blog/index.php/2019/08/14/a-crash-course-in-good-and-bad-control/

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

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

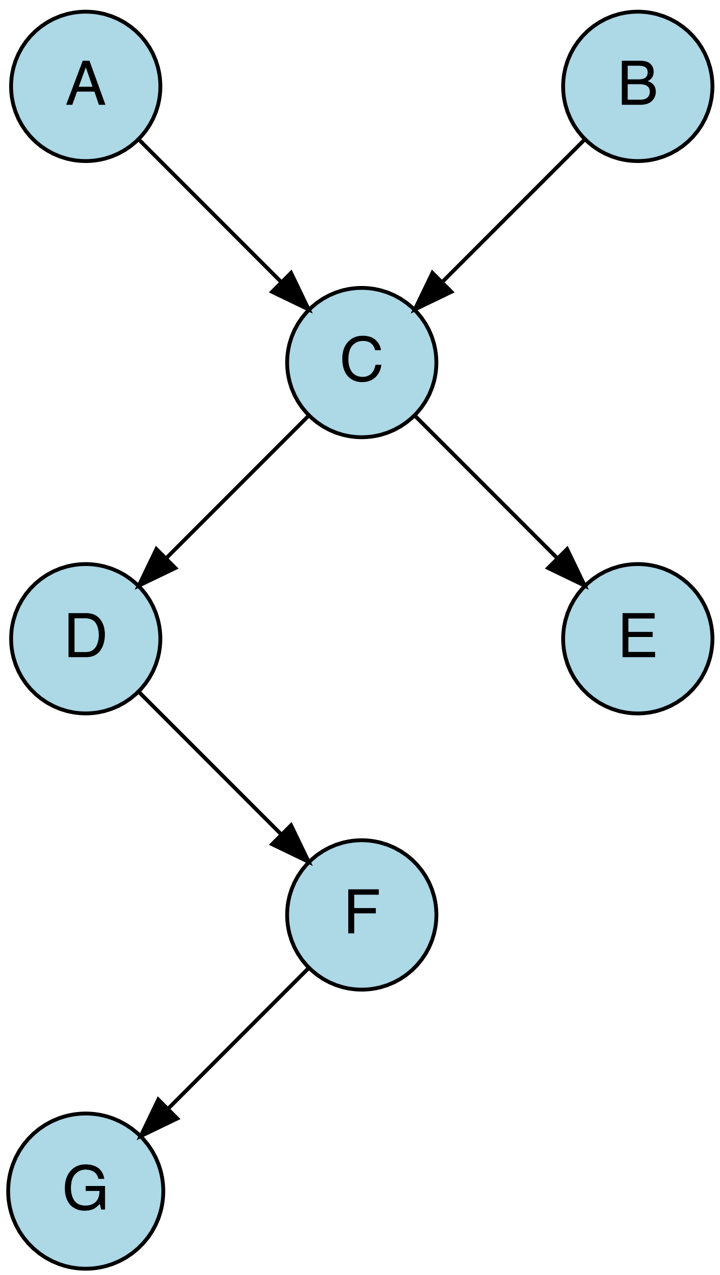
   - if variables are **disconnected** they are independent
   - if variables are connected (have a path between them) they are not guaranteed to be independent

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

When should I control for variables?

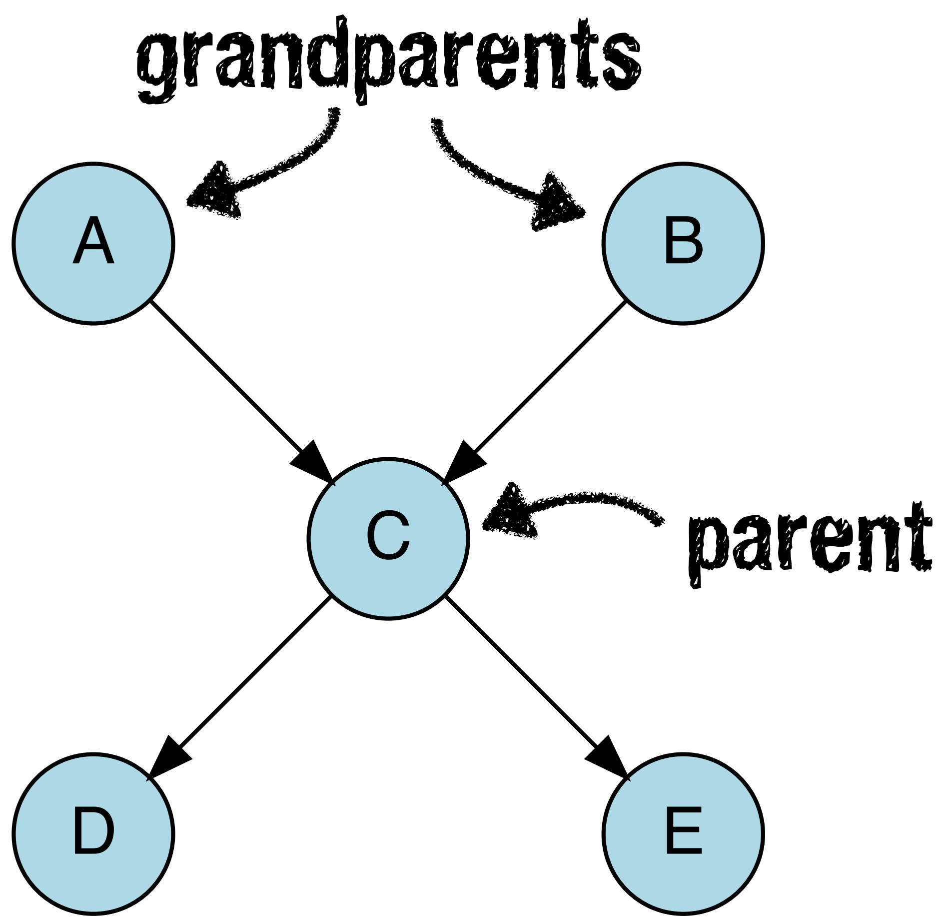
Are D and E independent?

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



1. Draw the ancestral graph

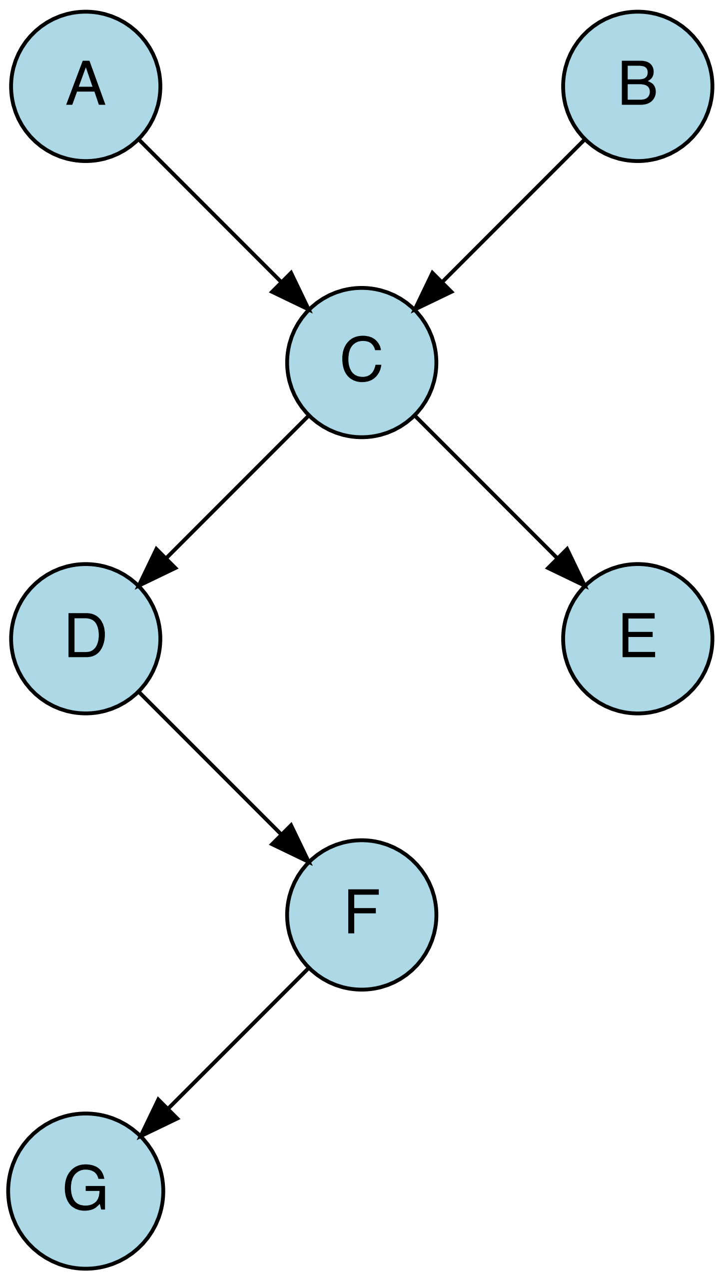
Construct the "ancestral graph" of all variables mentioned in the probability expression. This is a reduced version of the original net, consisting only of the variables mentioned and all of their ancestors (parents, parents' parents, etc.)



When should I control for variables?

Are D and E independent?

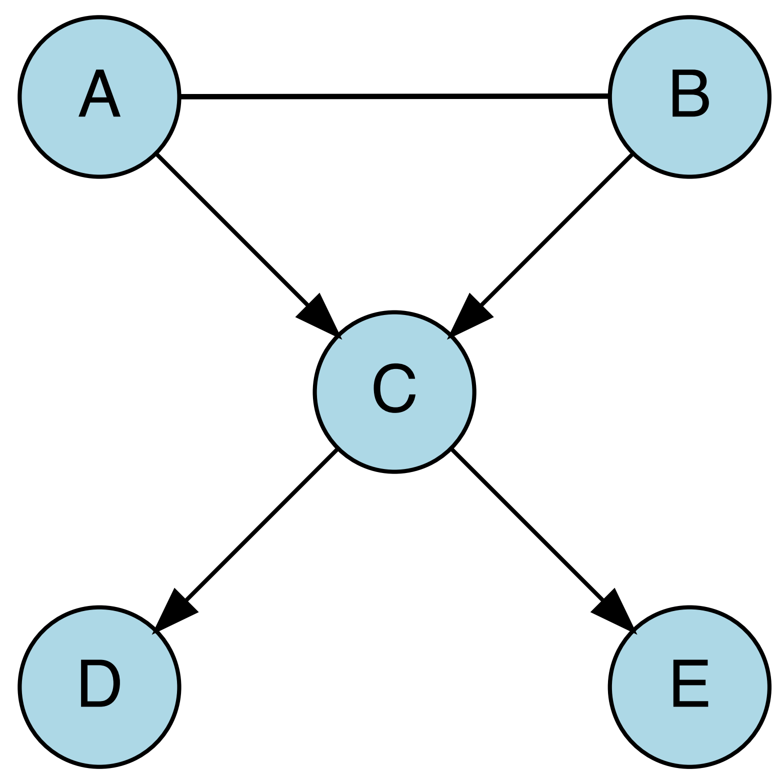
$$p(D \mid E) = p(D) \ ?$$



2. "Moralize" the graph

let's get married!

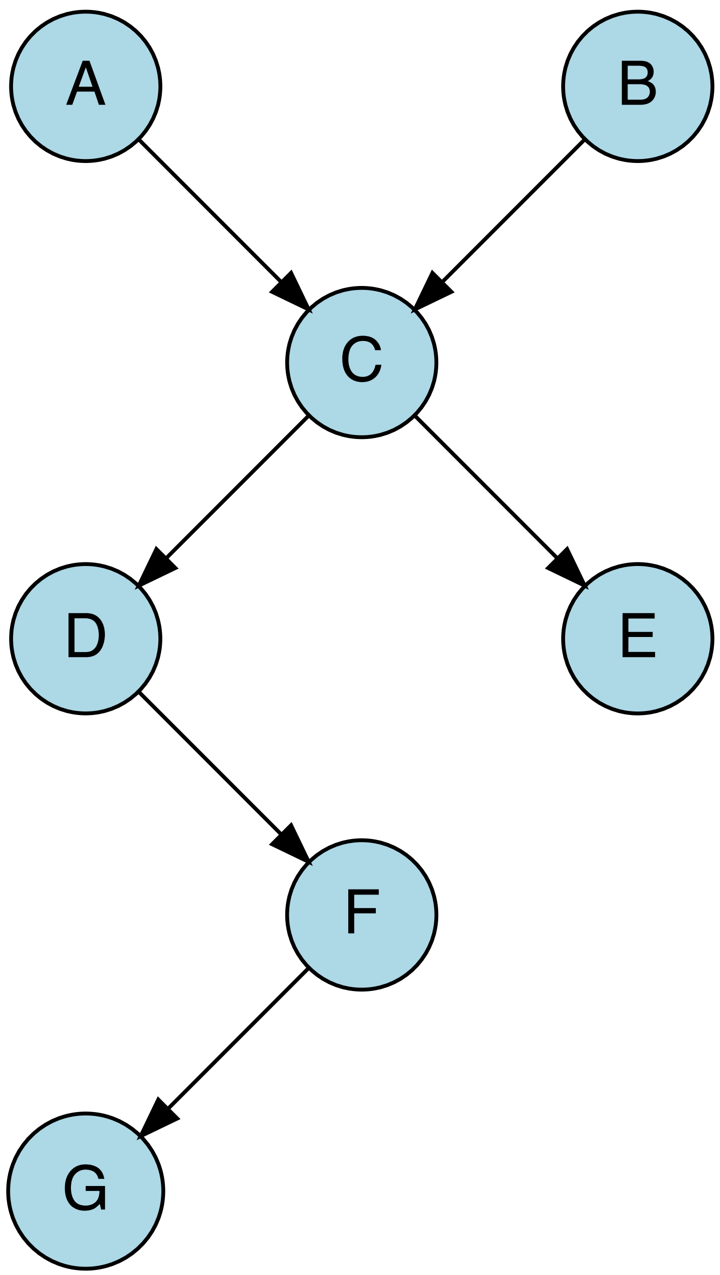
For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)



When should I control for variables?

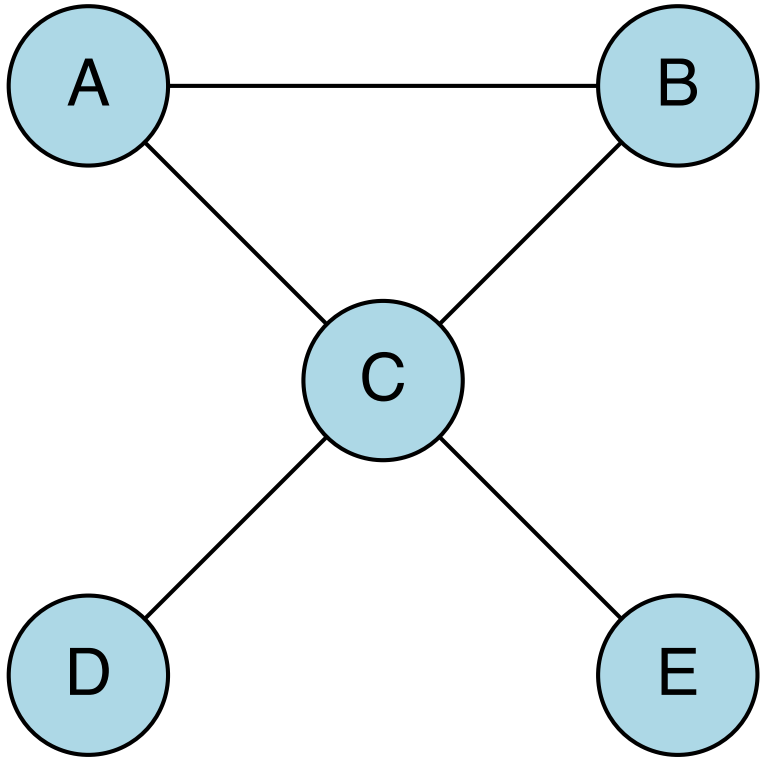
Are D and E independent?

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



3. "Disorient" the graph

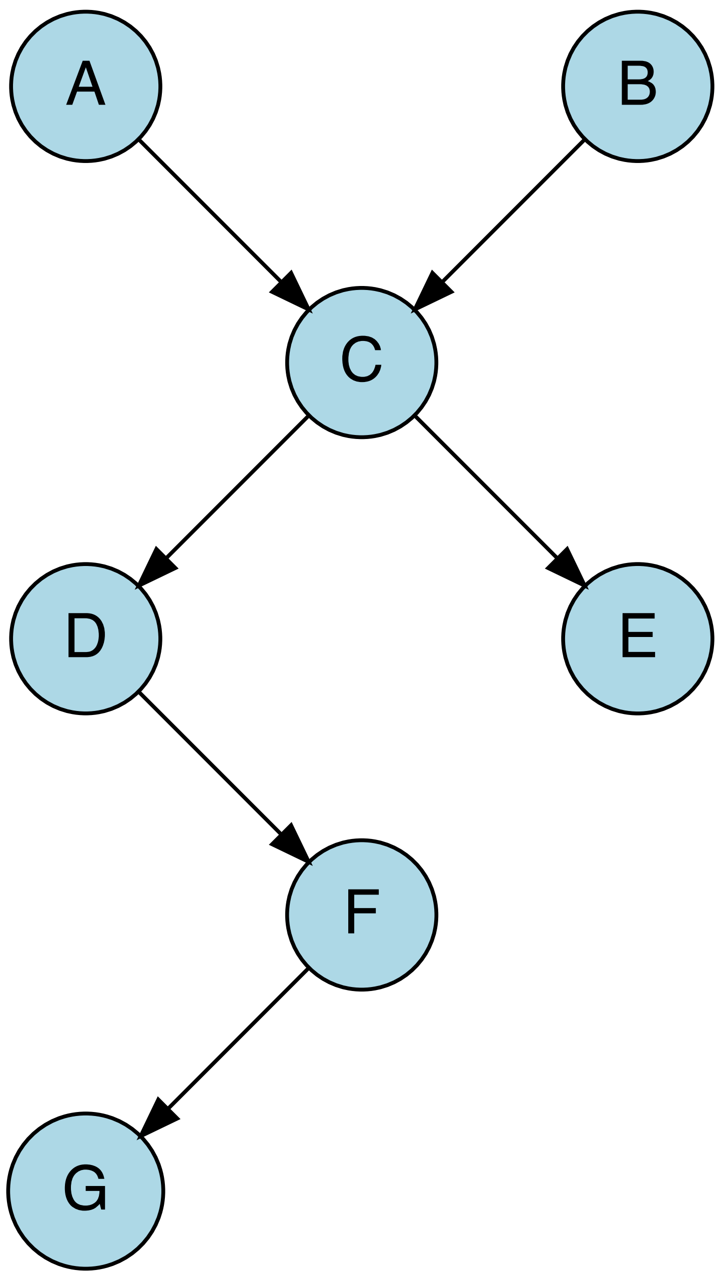
Replace arrows arrows with lines



When should I control for variables?

Are D and E independent?

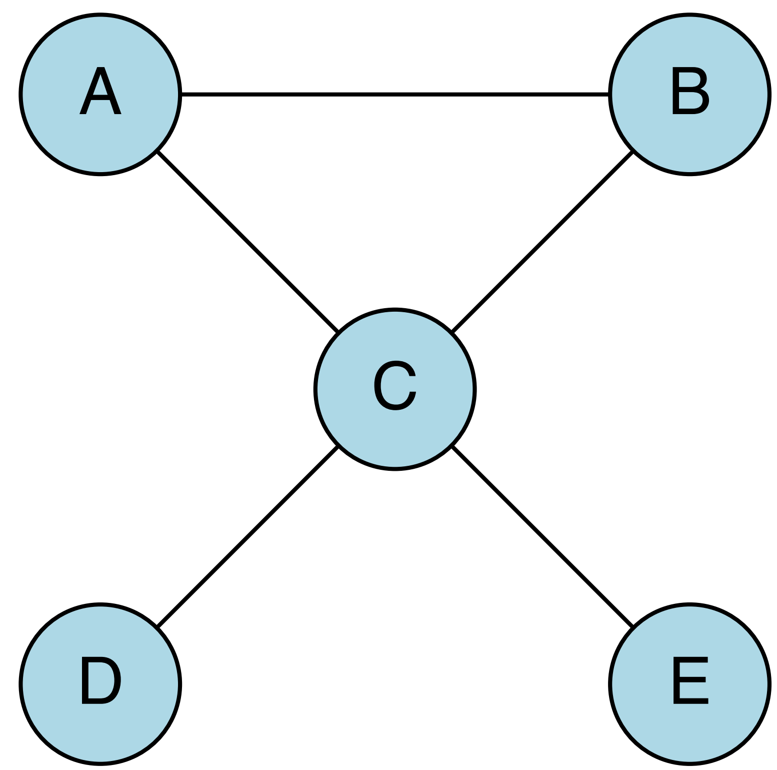
$$p(D \mid E) = p(D) \ ?$$



4. Delete the givens

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

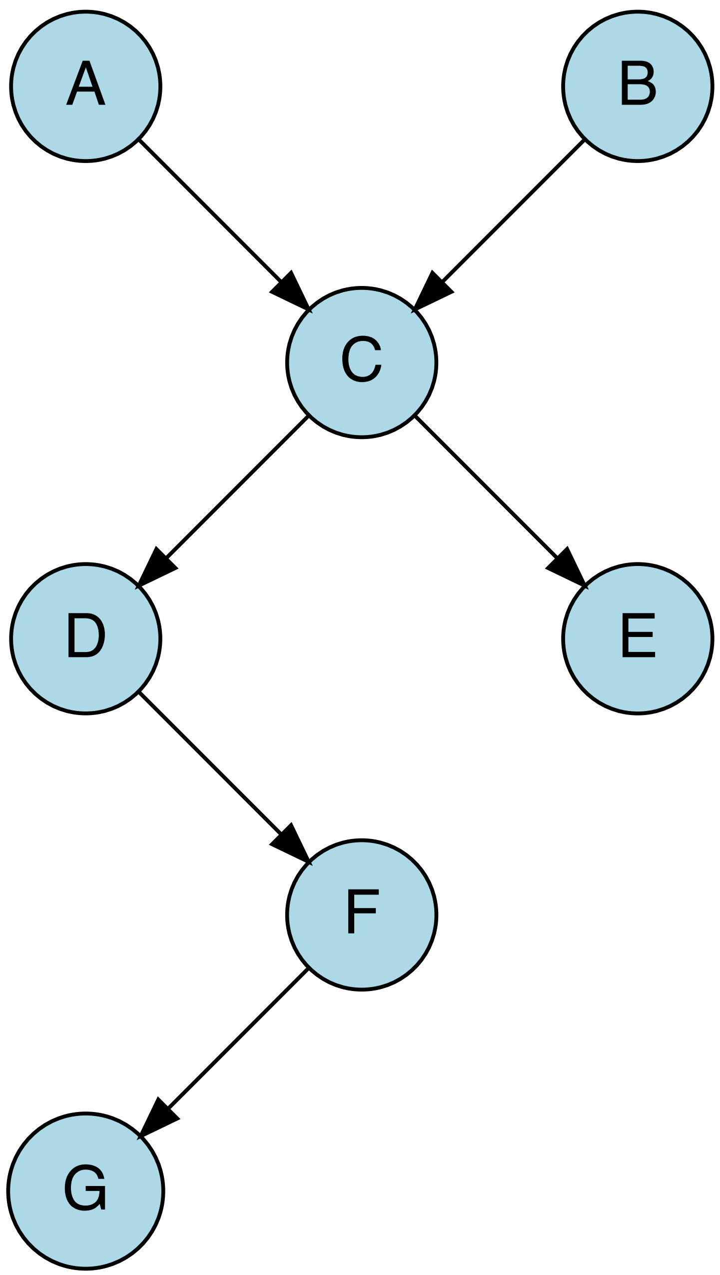
we didn't condition on anything, so there is nothing to delete



When should I control for variables?

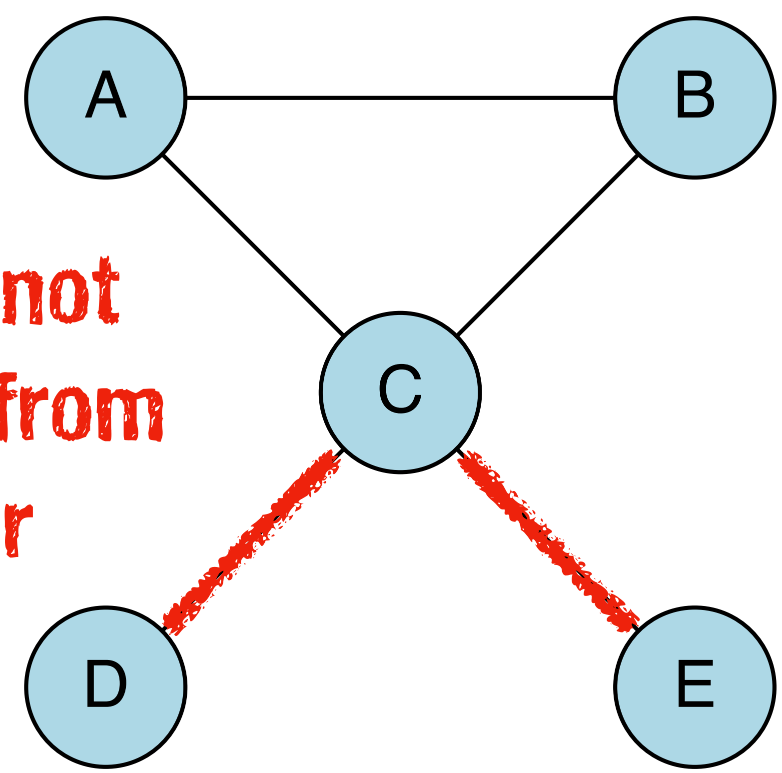
Are D and E independent?

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



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



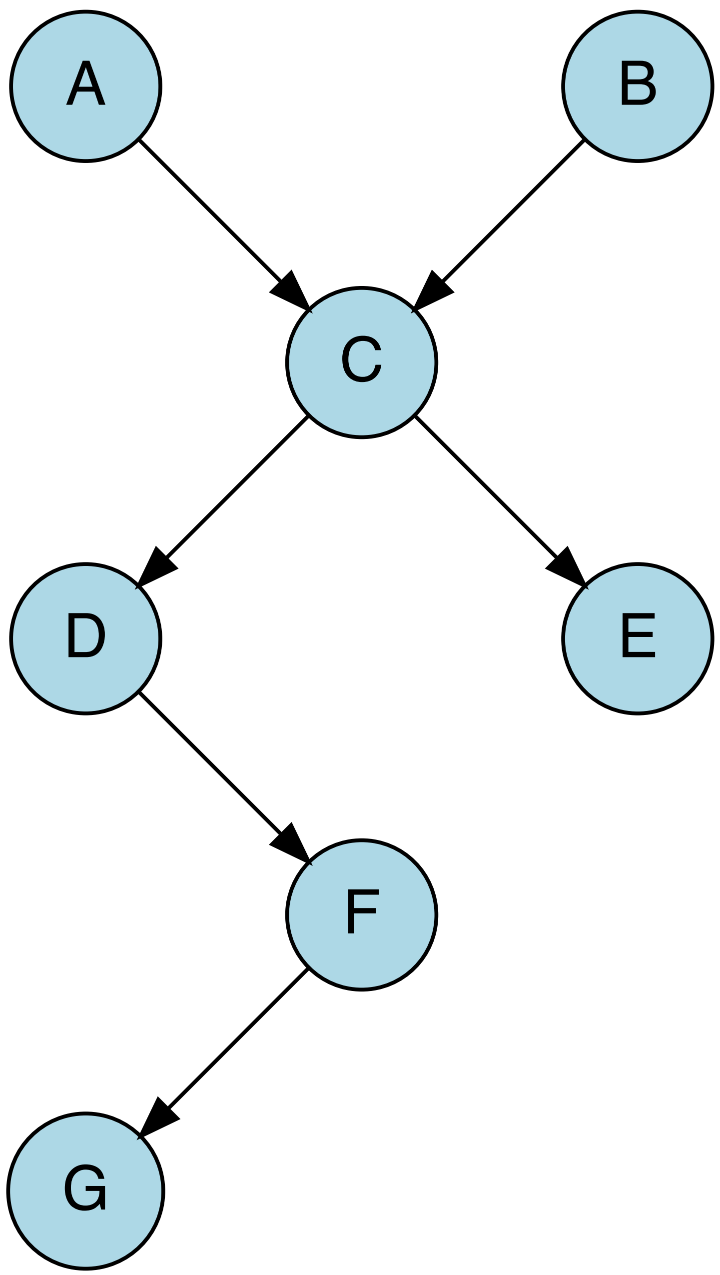
D and E are not independent from each other

they are connected via at least one path

When should I control for variables?

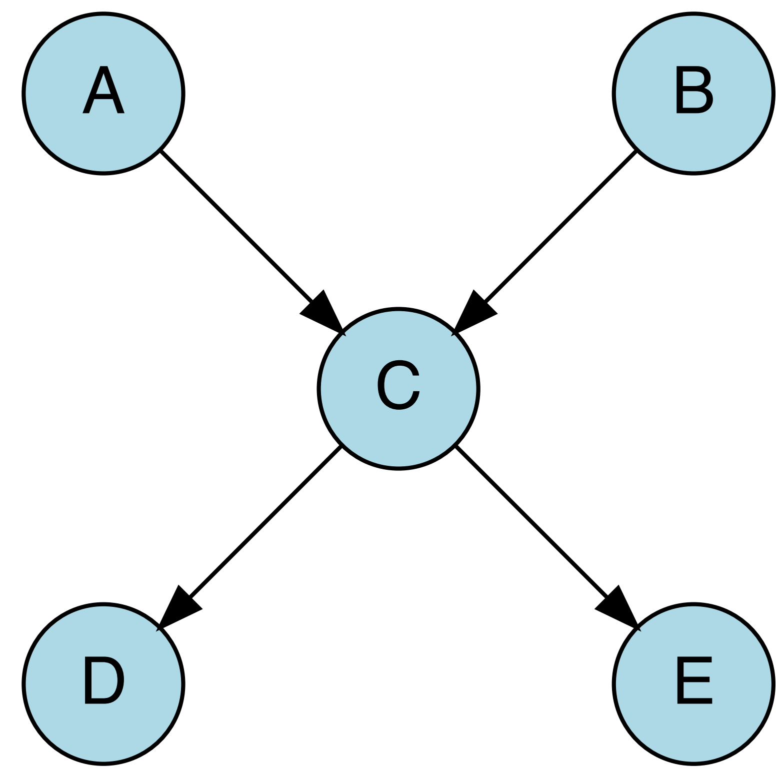
Are D and E independent, given C?

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



1. Draw the ancestral graph

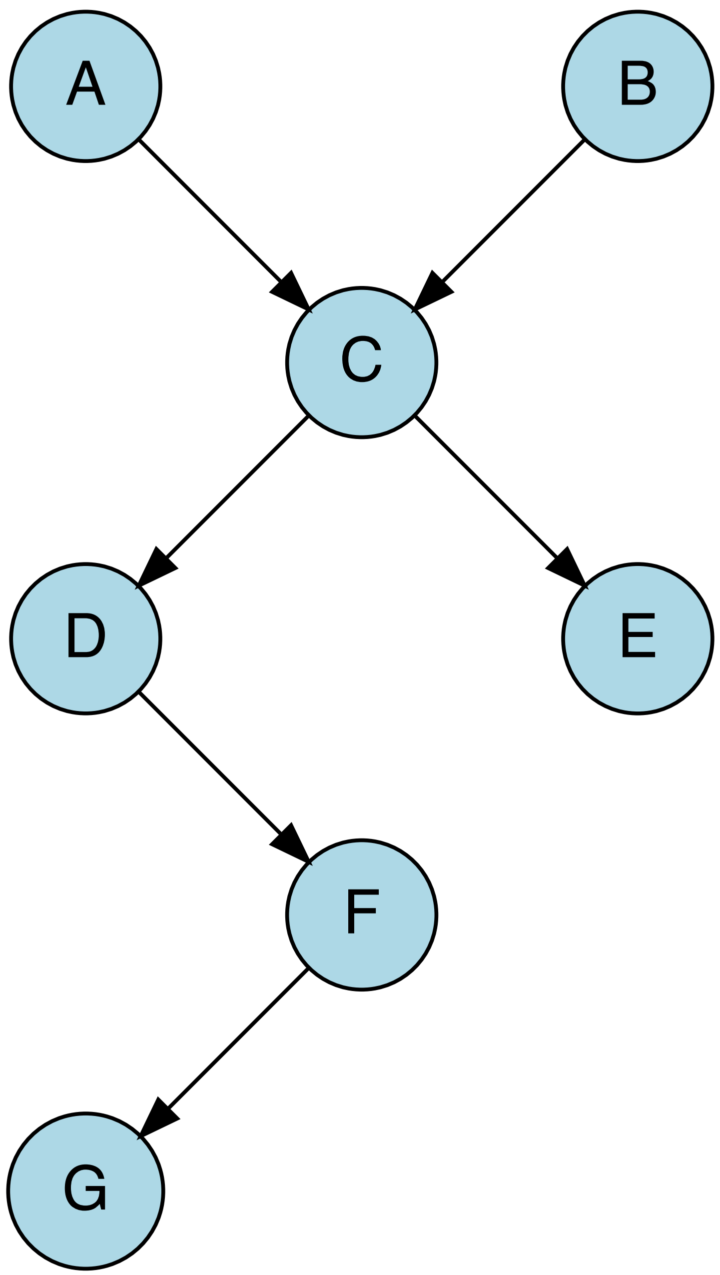
Construct the "ancestral graph" of all variables mentioned in the probability expression. This is a reduced version of the original net, consisting only of the variables mentioned and all of their ancestors (parents, parents' parents, etc.)



When should I control for variables?

Are D and E independent, given C?

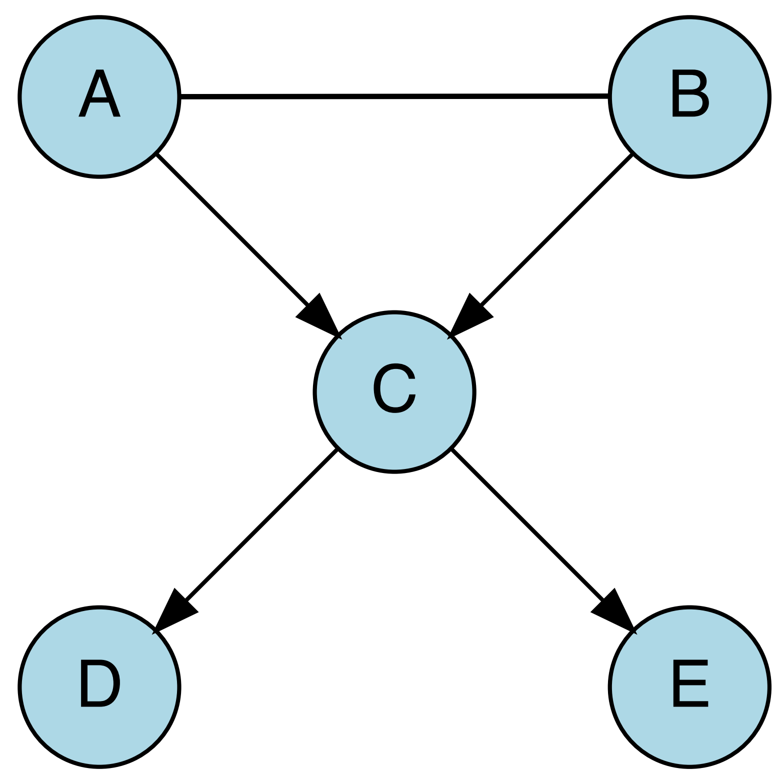
$$p(D \mid E, C) = p(D \mid C) \ ?$$



2. "Moralize" the graph

let's get married!

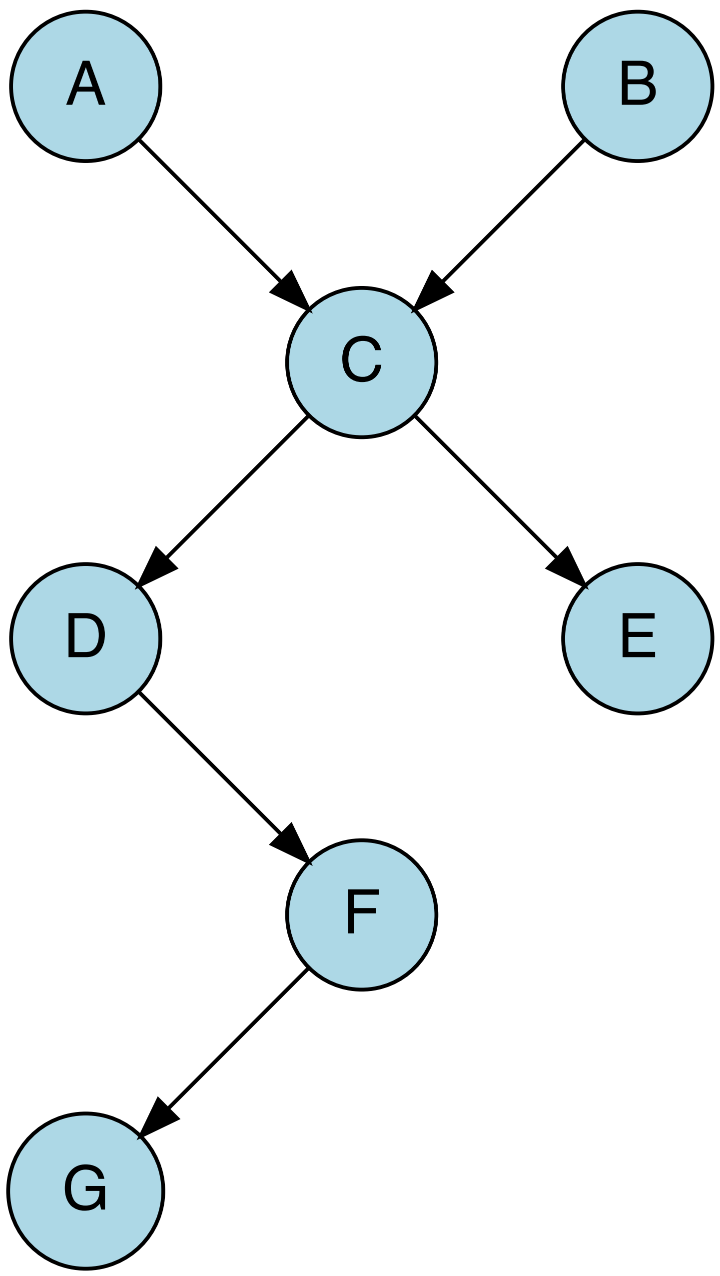
For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)



When should I control for variables?

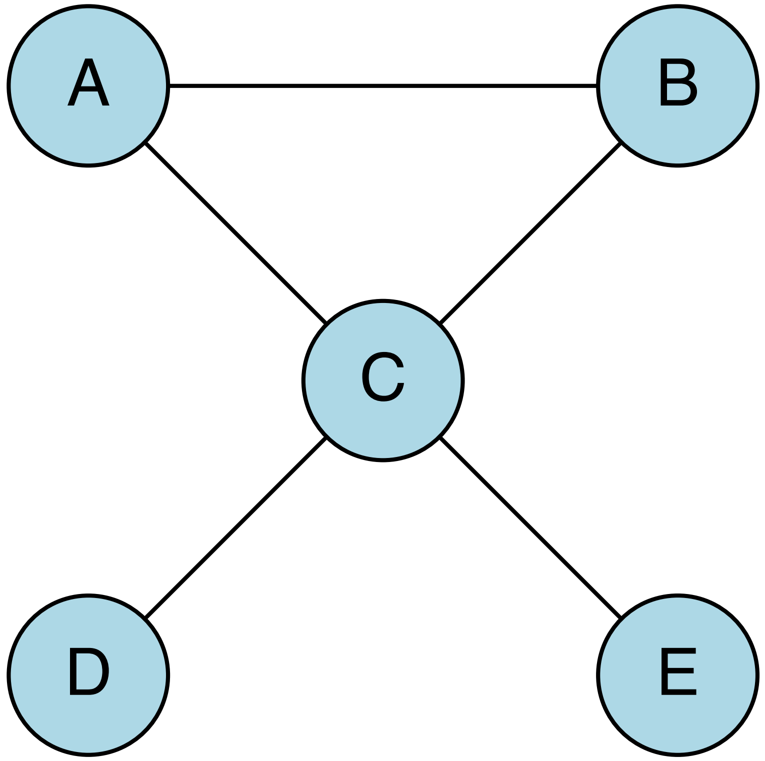
Are D and E independent, given C?

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



3. "Disorient" the graph

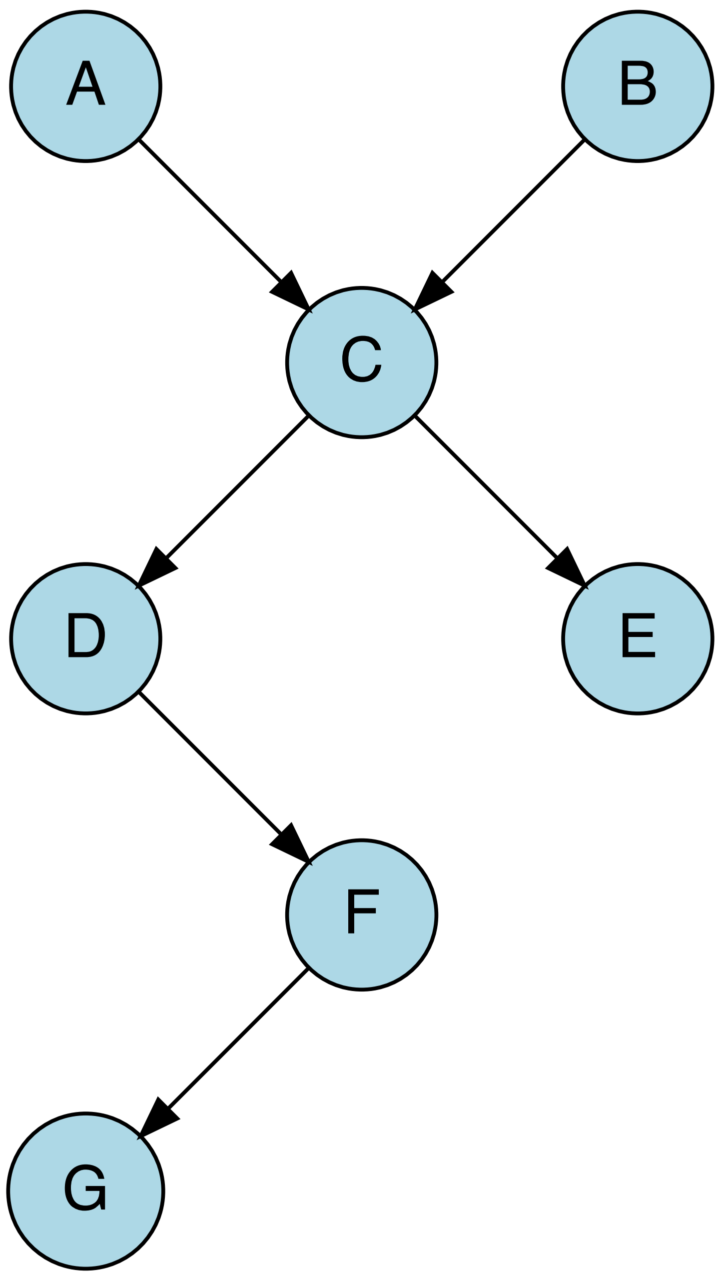
Replace arrows arrows with lines



When should I control for variables?

Are D and E independent, given C?

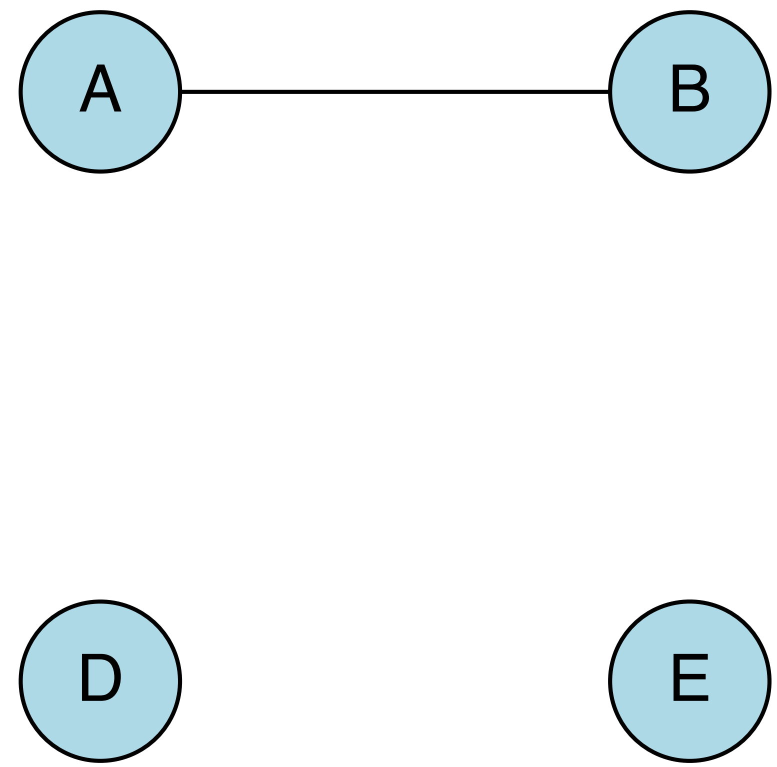
$$p(D \mid E, C) = p(D \mid C) \ ?$$



4. Delete the givens

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

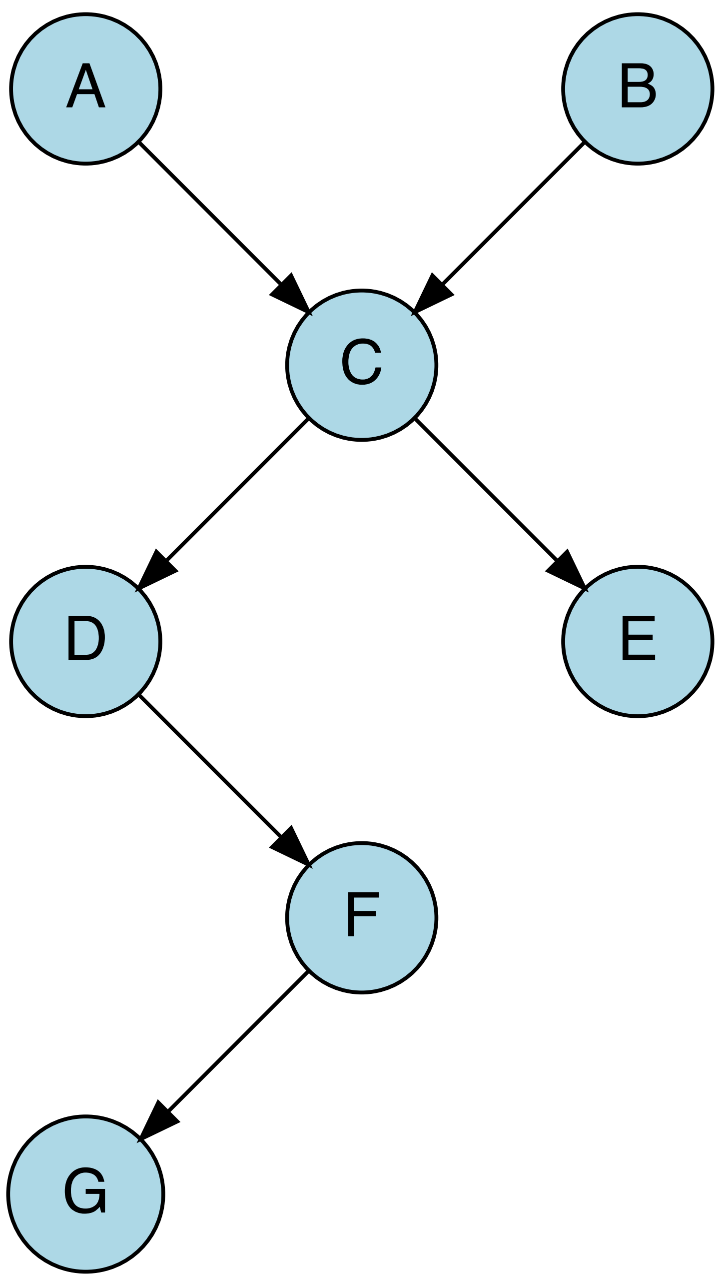
we conditioned on C!



When should I control for variables?

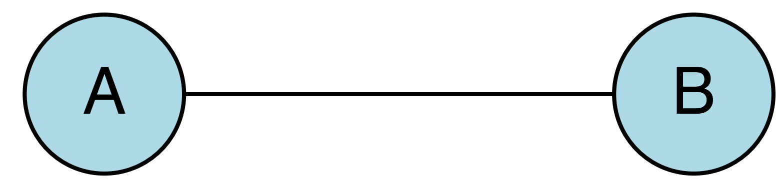
Are D and E independent, given C?

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



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



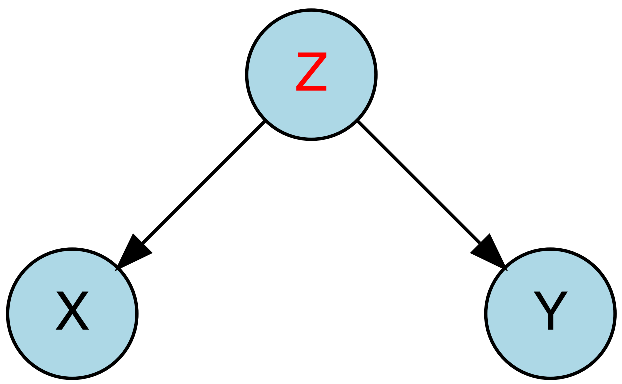
D and E are independent from each other conditioned on C

they aren't connected via a path

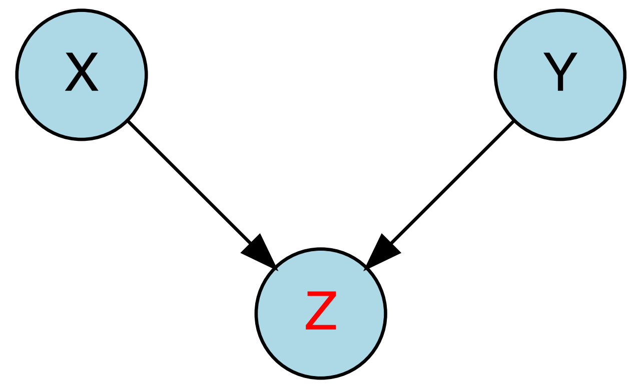
So what?

Patterns of inference

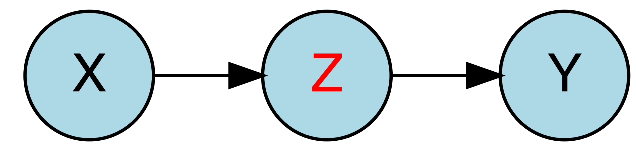
We want to estimate the (causal) relationship between `X` and `Y`



common cause



common effect



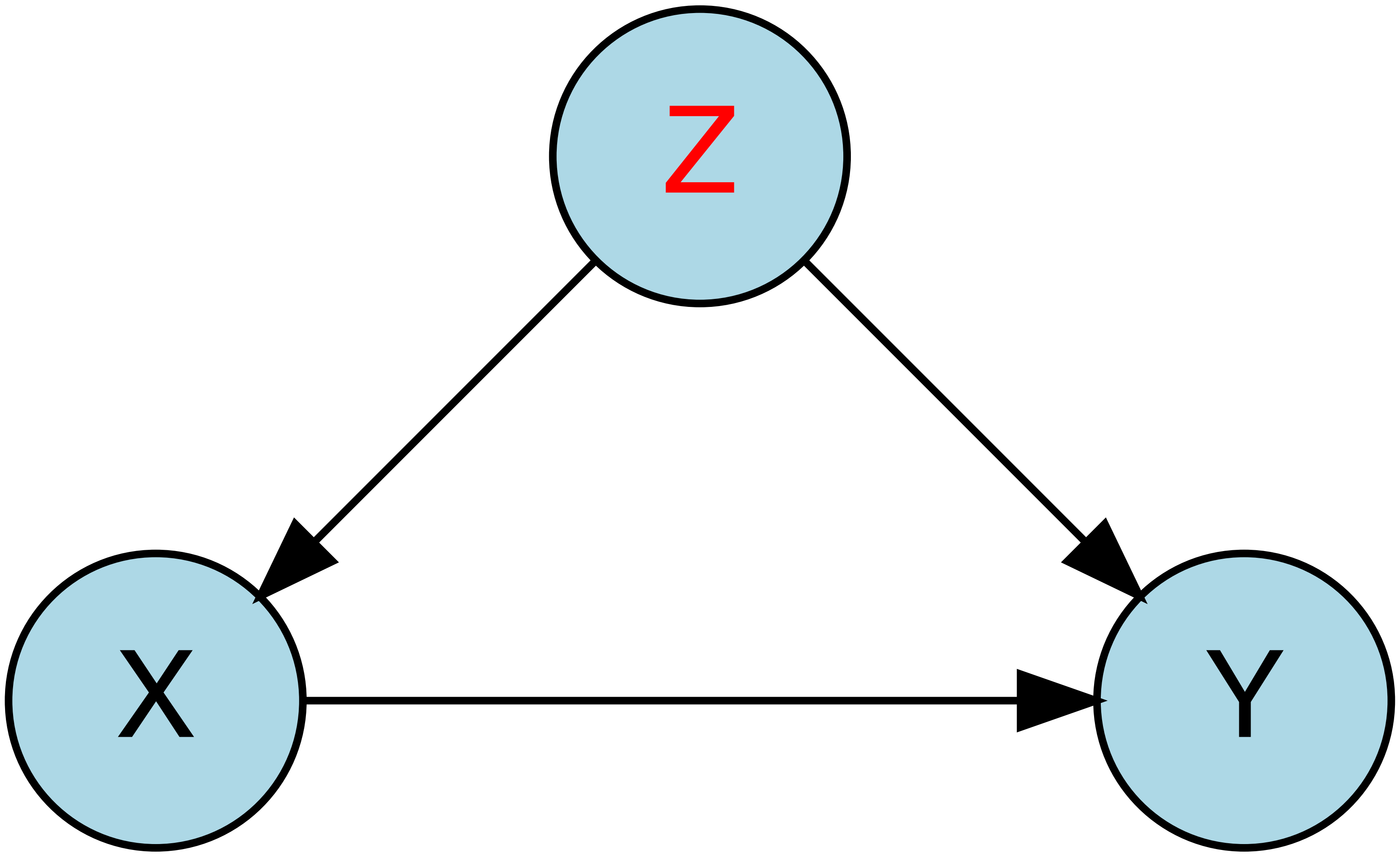
causal chain

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?

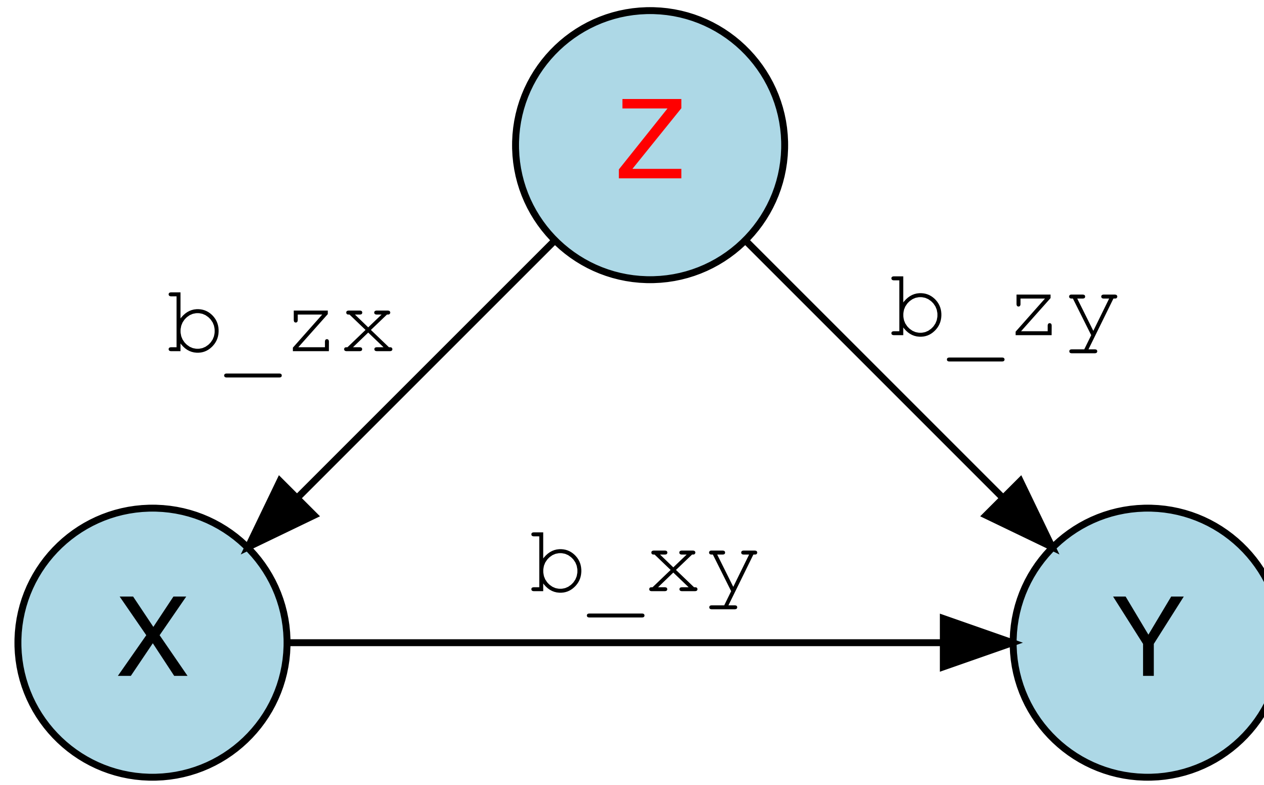
I want to estimate the effect that X has on Y



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

fun_error = function(n, sd){
  rnorm(n = n,
        mean = 0,
        sd = sd)
}

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



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

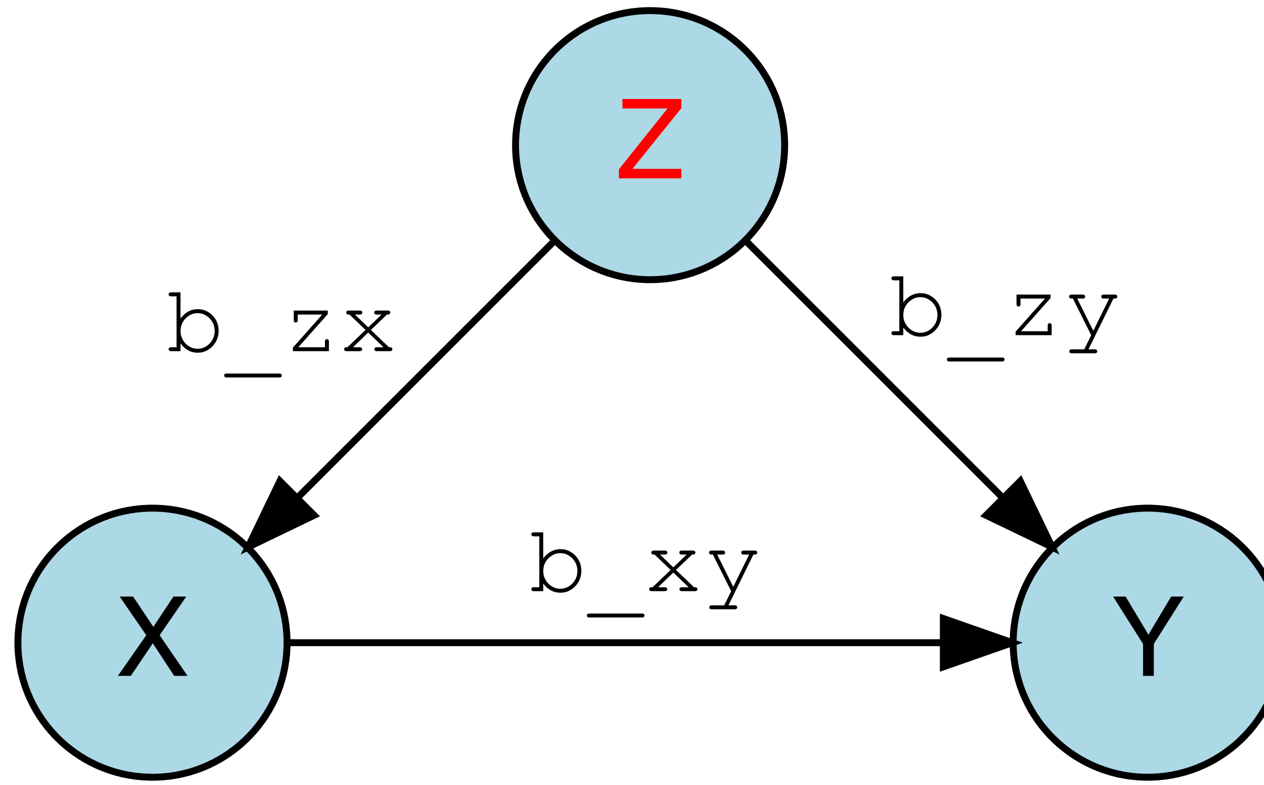
```r
# without control
lm(formula = y ~ x,
   data = df) %>%
  summary()
```

overestimating X's effect on Y

```
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

fun_error = function(n, sd){
  rnorm(n = n,
        mean = 0,
        sd = sd)
}

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



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

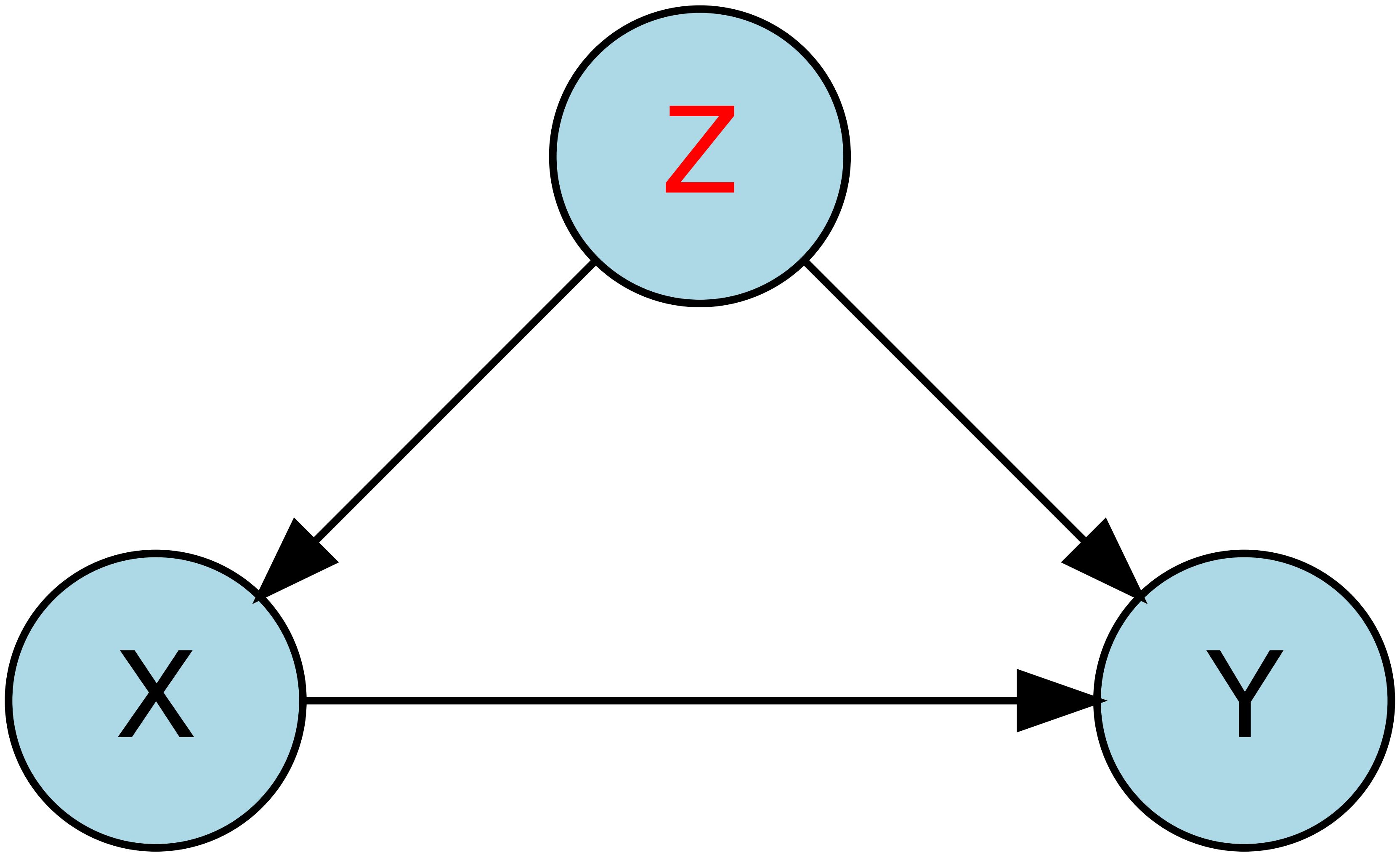
```r
# with control
lm(formula = y ~ x + z,
   data = df) %>%
  summary()
```

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

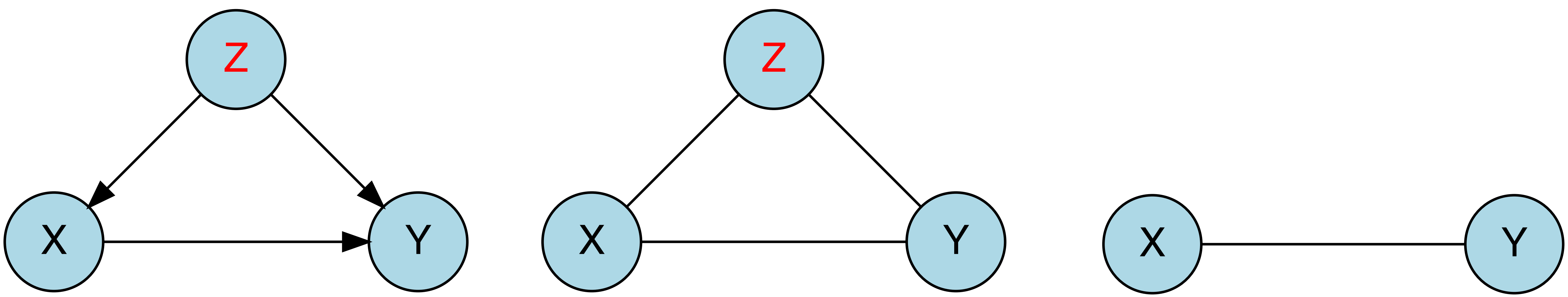


Z is a **good** control here!

When should I control for variables?

I want to estimate the effect that X has on Y

not controlling for Z | controlling for Z

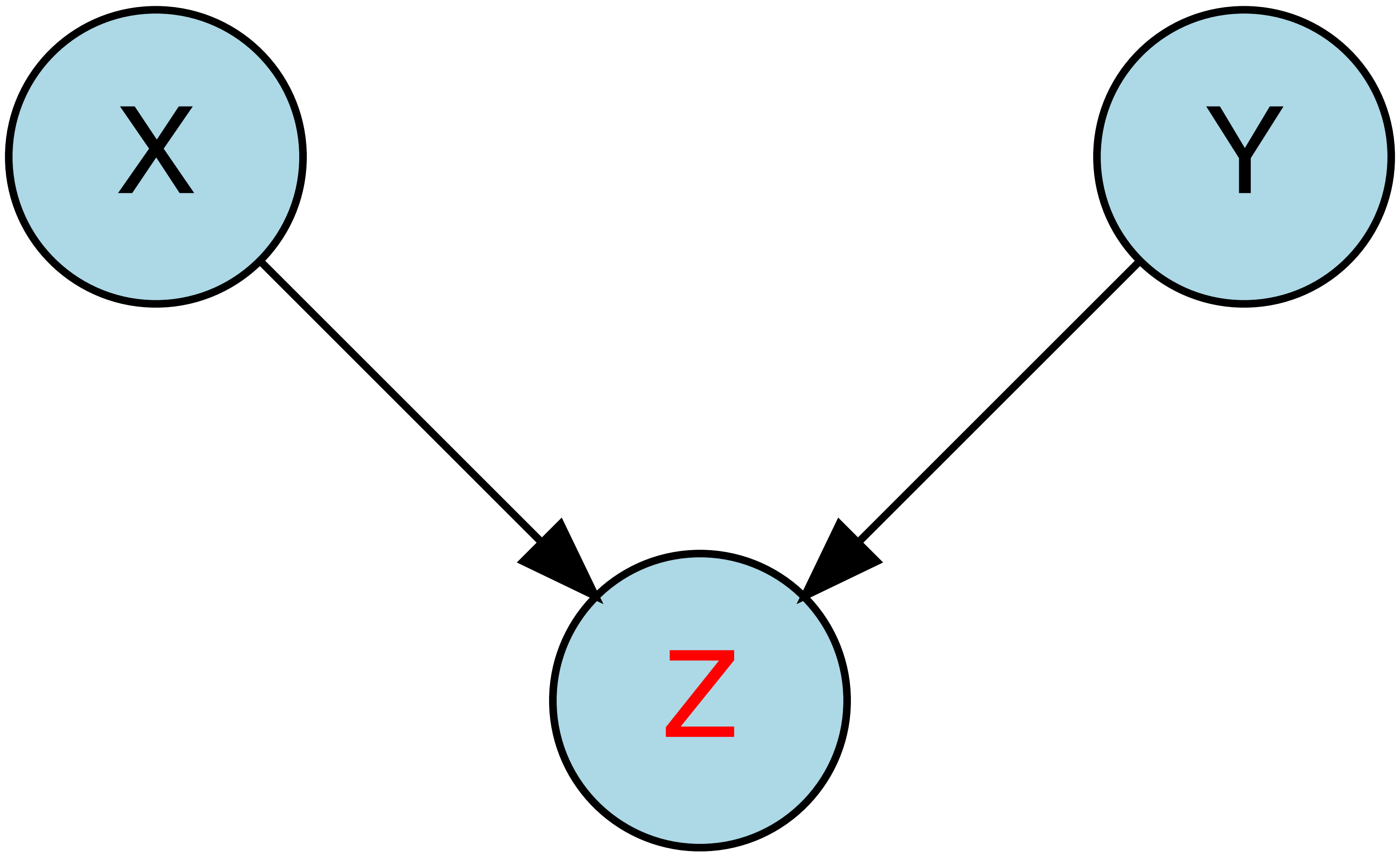


multiple paths | one path

Z is a **good** control here!

When should I control for variables?

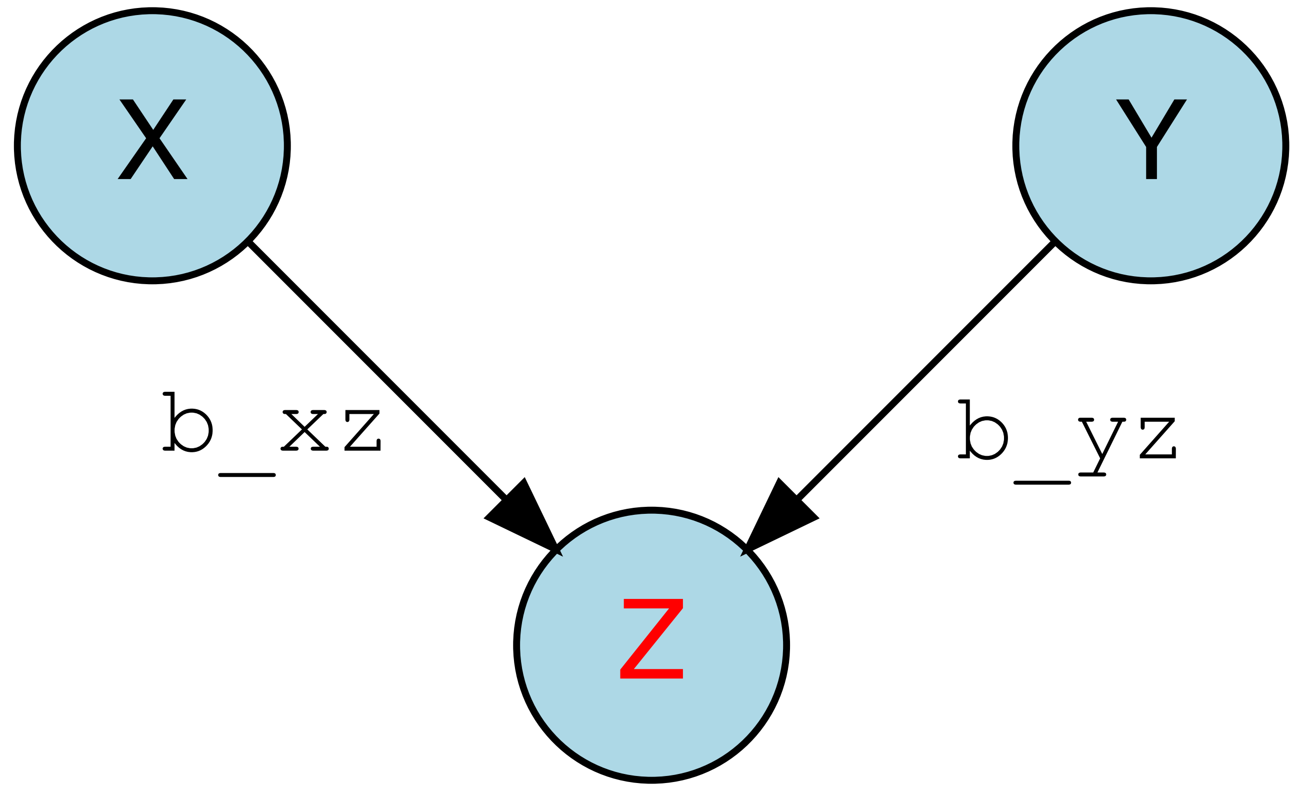
I want to estimate the effect that X has on Y



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

fun_error = function(n, sd){
  rnorm(n = n,
        mean = 0,
        sd = sd)
}

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



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

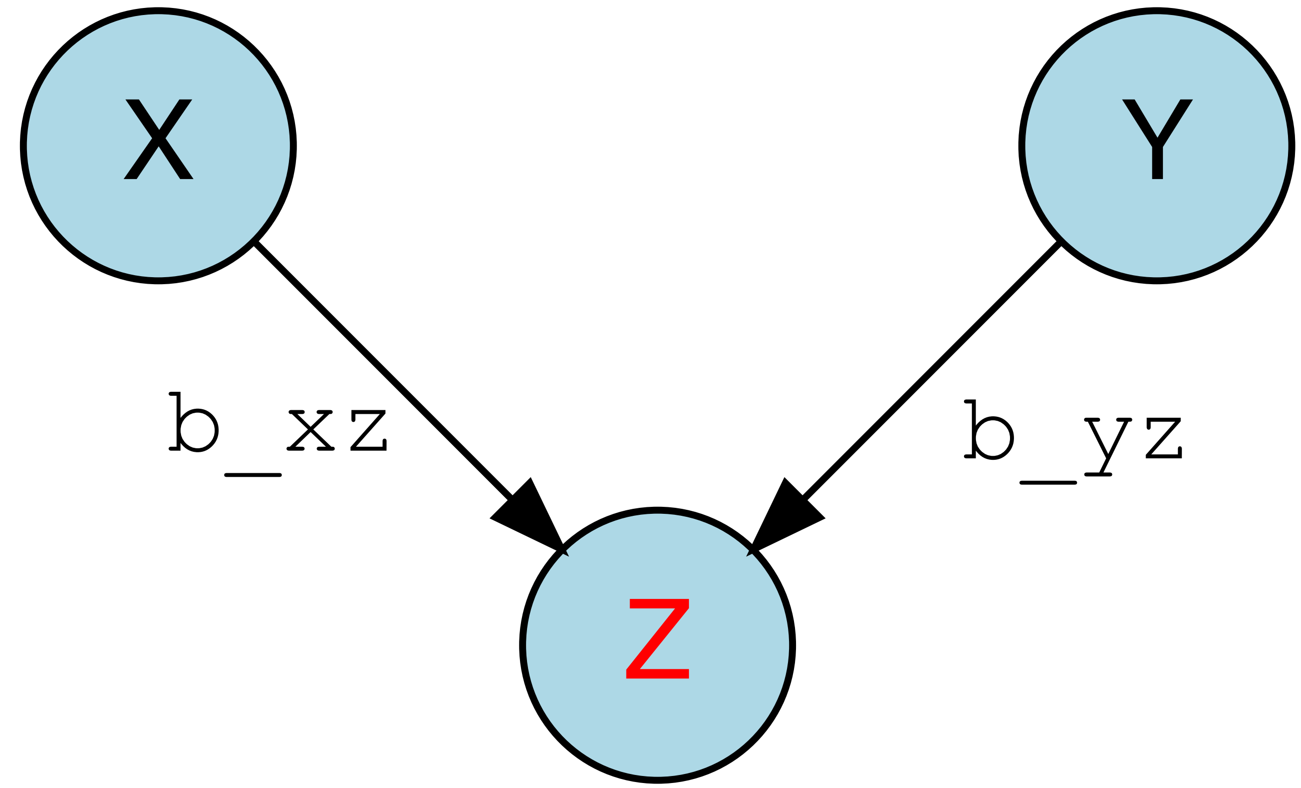
```r
# without control
lm(formula = y ~ x,
   data = df) %>%
  summary()
```

accurate estimate of X's effect on Y

```
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

fun_error = function(n, sd){
  rnorm(n = n,
        mean = 0,
        sd = sd)
}

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



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

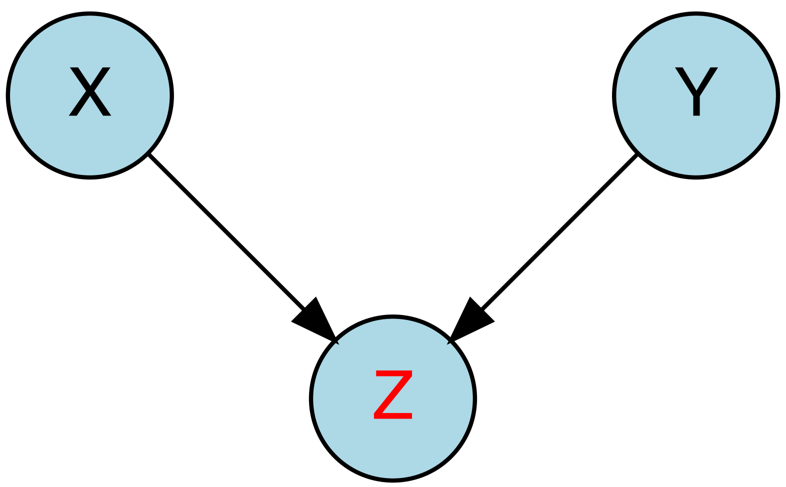
```r
# with control
lm(formula = y ~ x + z,
   data = df) %>%
  summary()
```

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?

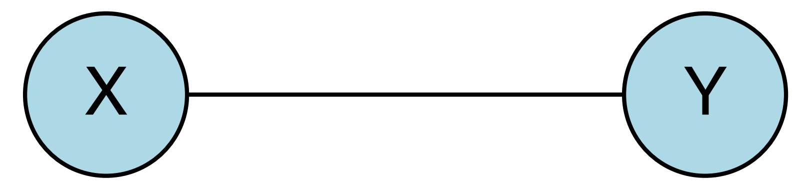
I want to estimate the effect that X has on Y



not controlling for Z

no path

controlling for Z



one path

Z is a **bad** control here!

When should I control for variables?

Babies Learning Language

Thoughts on language learning, child development, and fatherhood; experimental methods, reproducibility, and open science; theoretical musings on cognitive science more broadly.

Tuesday, October 8, 2019

Confounds and covariates

(tl;dr: explanation of confounding and covariate adjustment)

Every year, one of the trickiest concepts for me to teach in my experimental methods course is the difference between experimental confounds and covariates. Although this distinction seems simple, it's pretty deeply related to the definition of what an experiment is and why experiments lead to good causal inferences. It's also caught up in a number of methodological problems that come up again and again in my class. This post is my attempt to explain the distinction and how it relates to different problems and cultural practices in psychology.

About

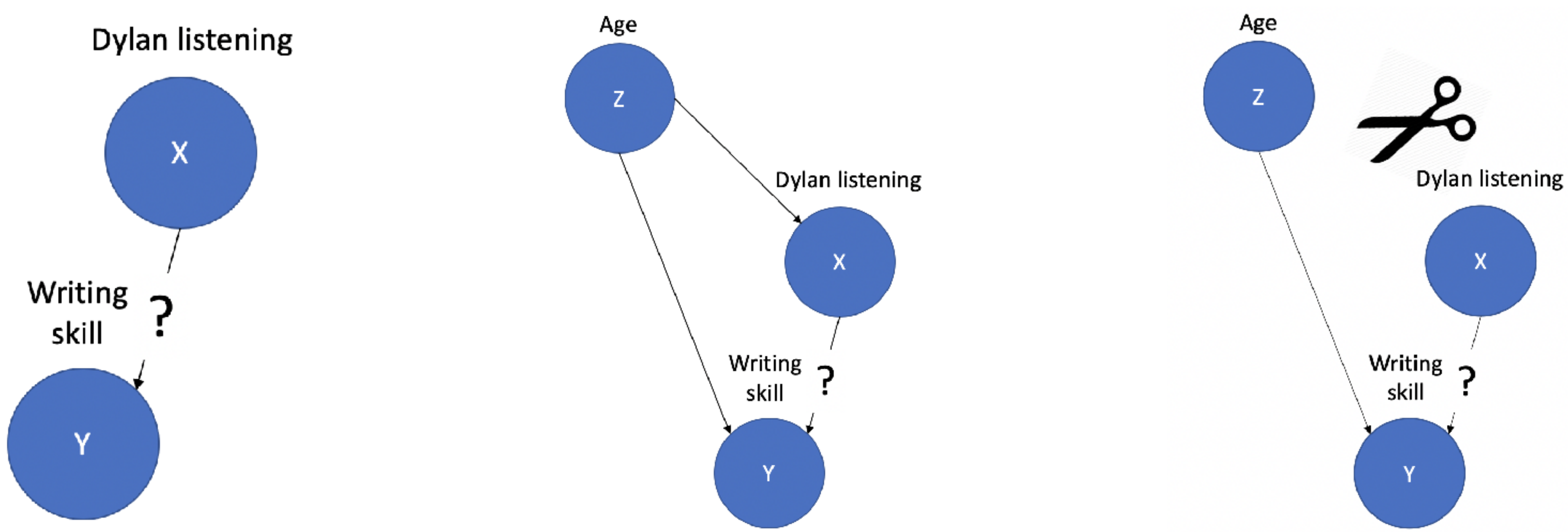
Michael Frank

Developmental Psychologist at Stanford University. Interested in language acquisition, cognitive development, thoughts, numbers, learning, and teaching. Visit my lab's website.

View my complete profile

Labels

- Development (40)



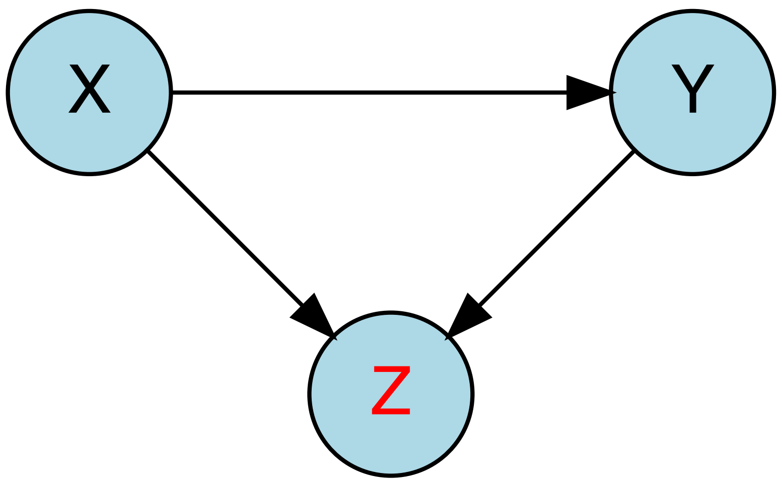
http://babieslearninglanguage.blogspot.com/2019/10/confounds-and-covariates.html

When should I control for variables?

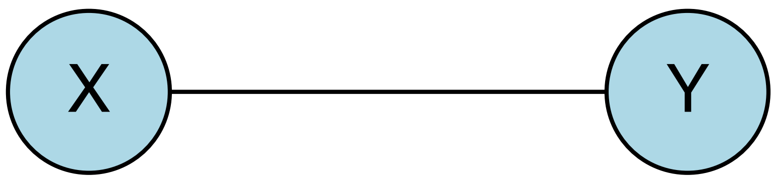
- checking for **d-separation** tells us whether or not variables are (conditionally) independent
- it also tells us whether paths of dependence "open up", or get "closed down"
- the graphical procedure doesn't necessarily reveal whether the dependence between variables changes: it reveals the **structure** 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

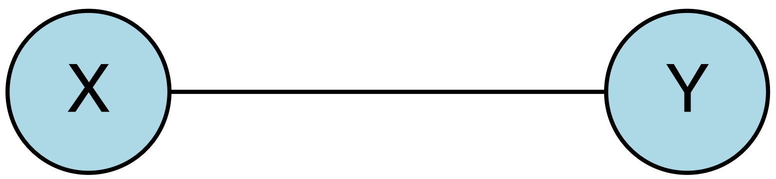


not controlling for Z



one path

controlling for Z



change in dependence

not revealed in the undirected graph

Z is a **bad** control here!

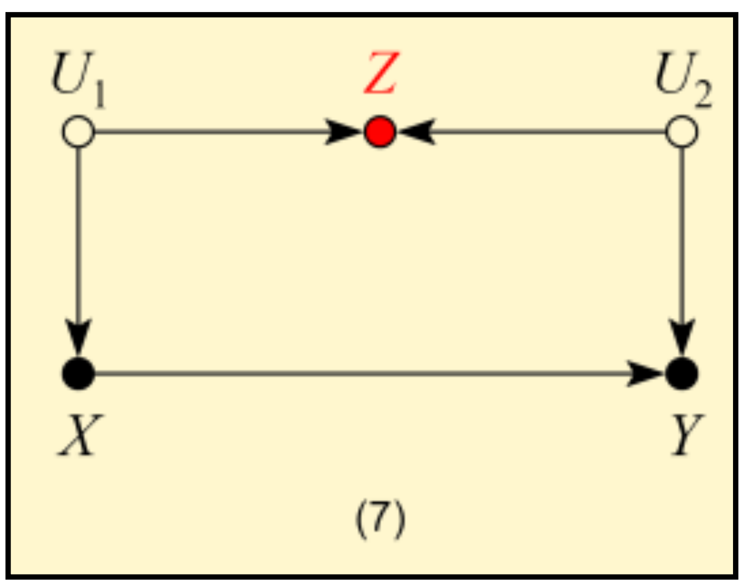
When should I control for variables?

- **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?

"Important topic. I feel like this procedure for determining good and bad controls is a bit convoluted. I feel like there's a simpler shorthand. Does another factor influence both x and y? Does another factor influence x? What's wrong with that simpler algorithm?"

not enough unfortunately



Z is a **bad** control here!

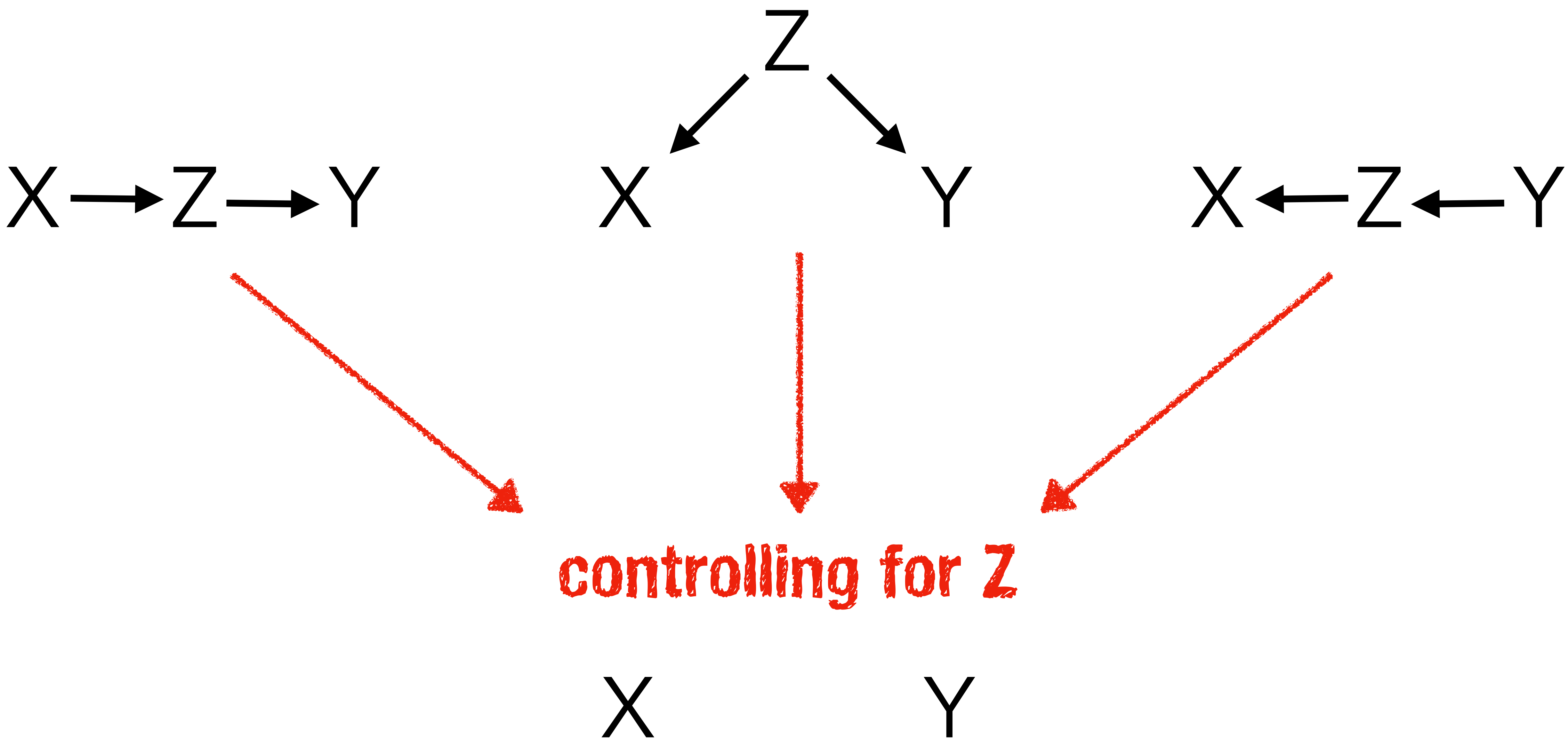
not controlling for Z

controlling for Z

additional path

When should I control for variables?

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

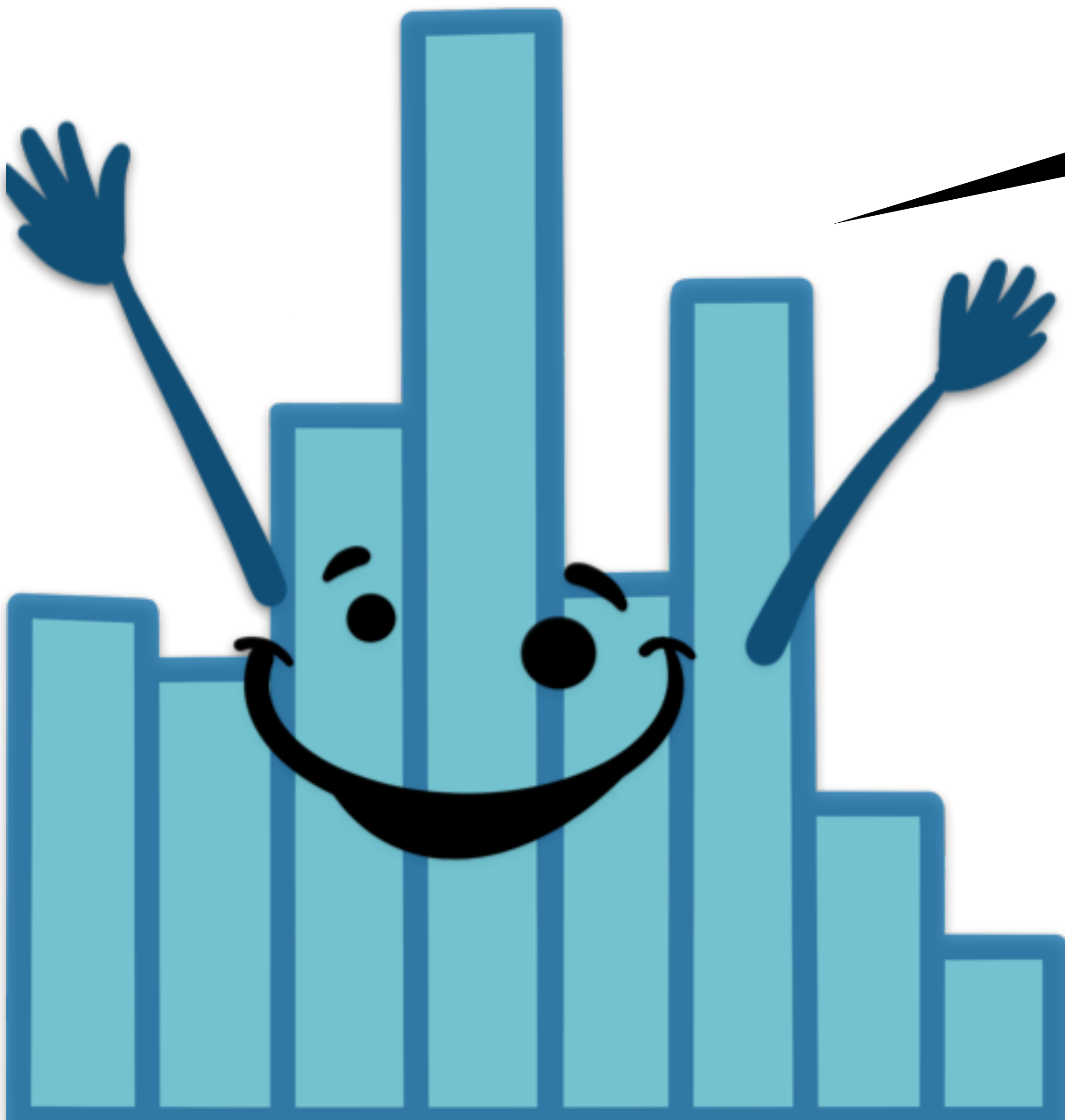


we need to manipulate X experimentally to tell these apart*

* sort of (see next slide)

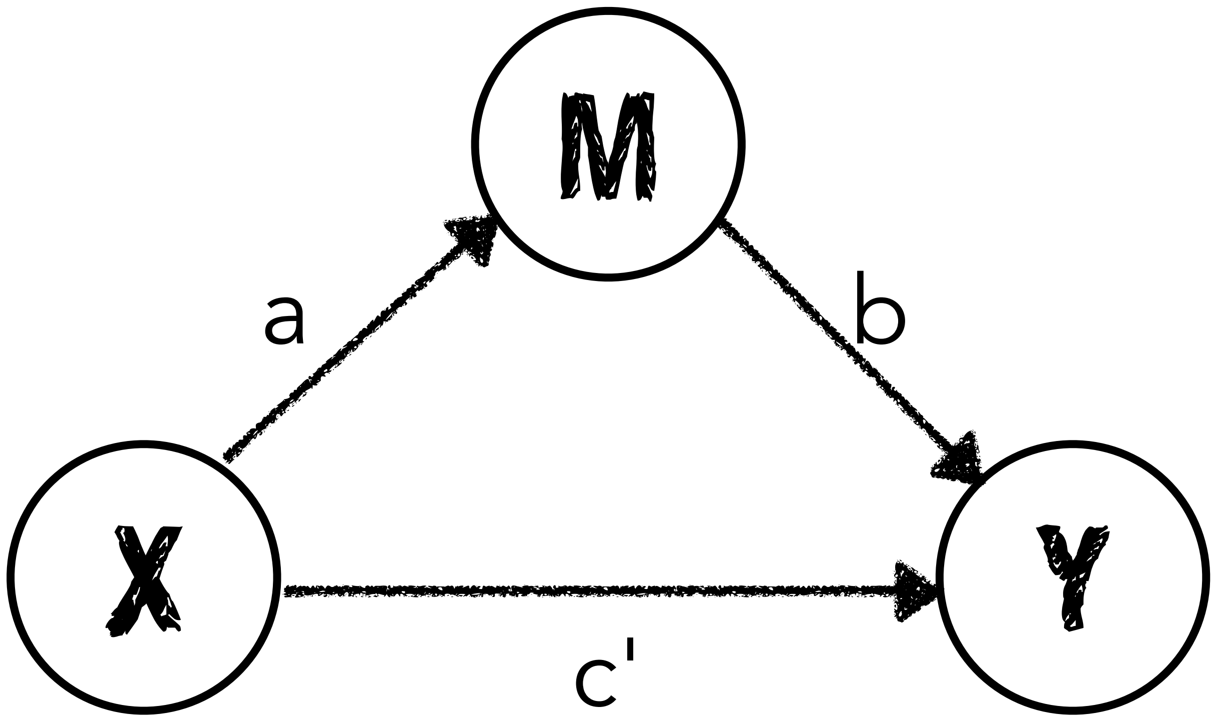
01:00

stretch break!



Mediation

Definition

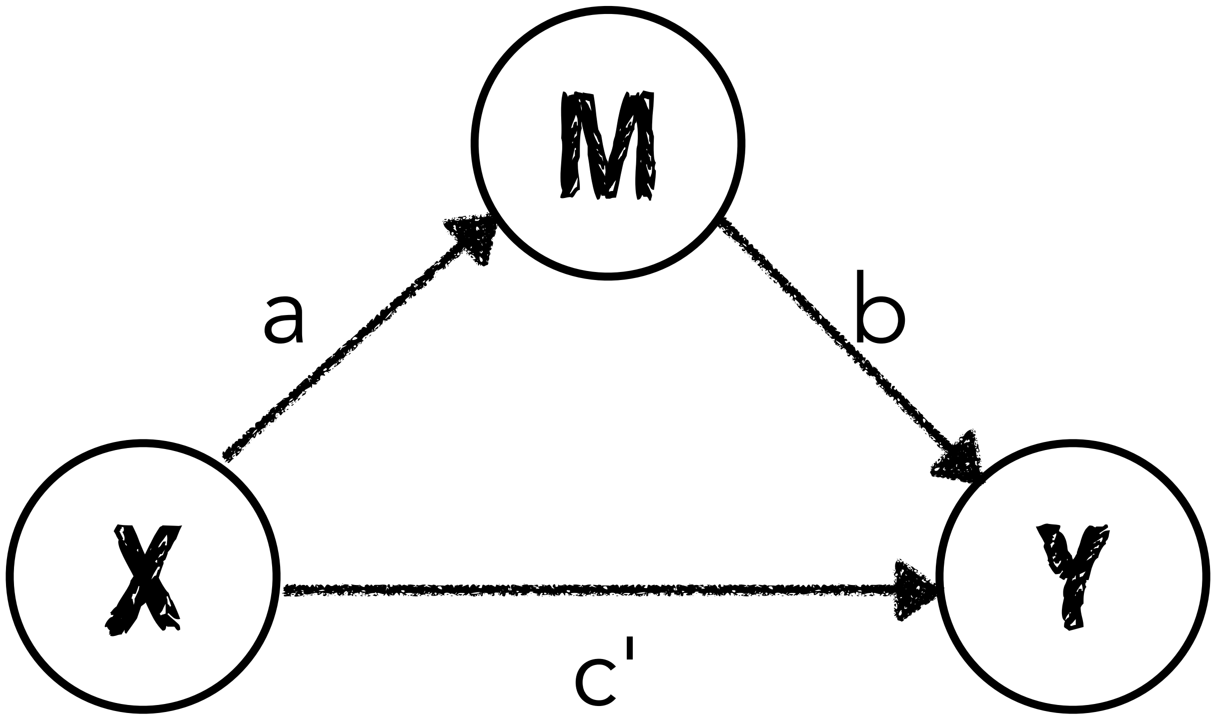


Rather than a direct causal relationship between **X** and **Y**, a mediation model proposes that **X** influences the mediator variable **M**, which in turn influences **Y**. Thus, the mediator variable serves to clarify the nature of the relationship between **X** and **Y**.

Adapted from Wikipedia

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

Example



X = grades in Psych 252

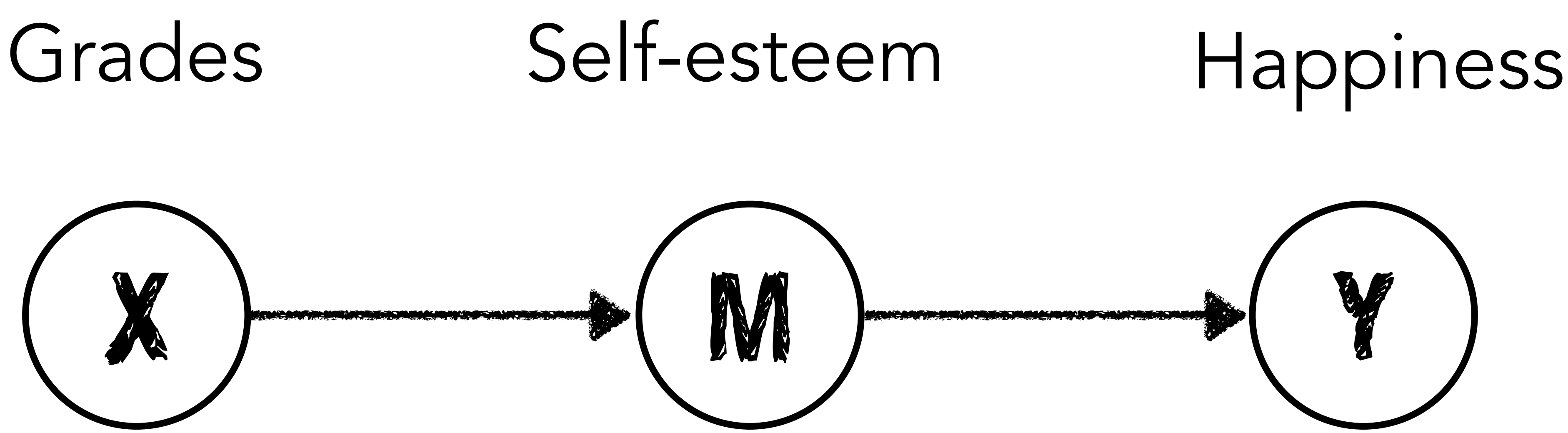
M = feelings of self-esteem

Y = happiness

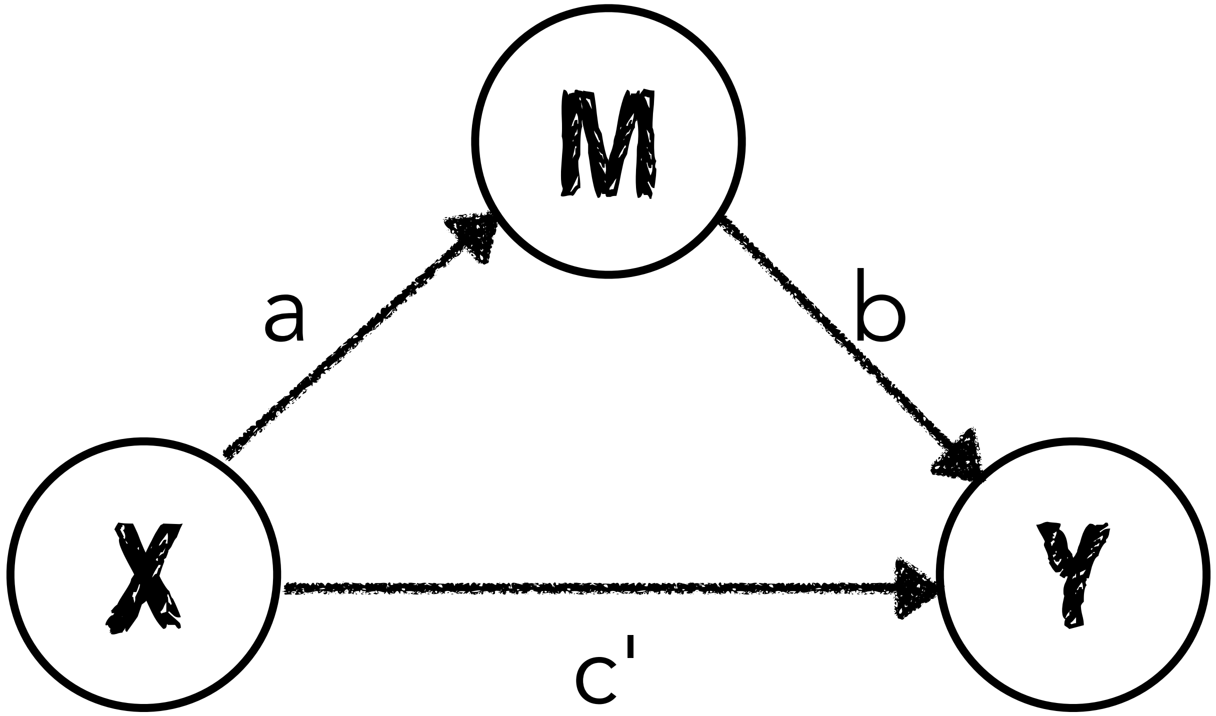
Is the relationship between grades in Psych 252 and happiness mediated by feelings of self-esteem?

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
)
```



Baron and Kenny's (1986) steps for mediation



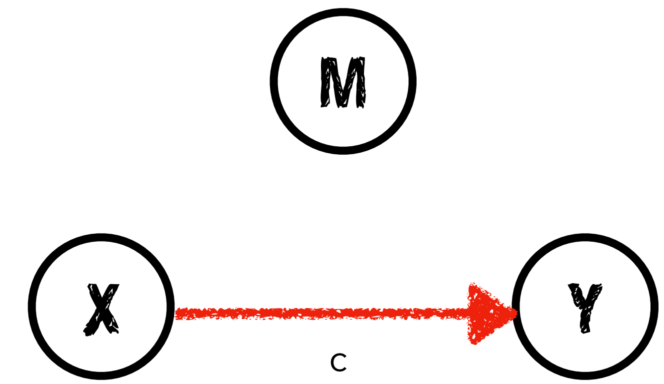
Sequence of regression models

1. Is there a relationship between **X** and **Y**?
2. Is there a relationship between **X** and **M**?
3. Does the relationship between **X** and **Y** change, once we control for **M**?

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.

Is there a relationship between **X** and **Y**?

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



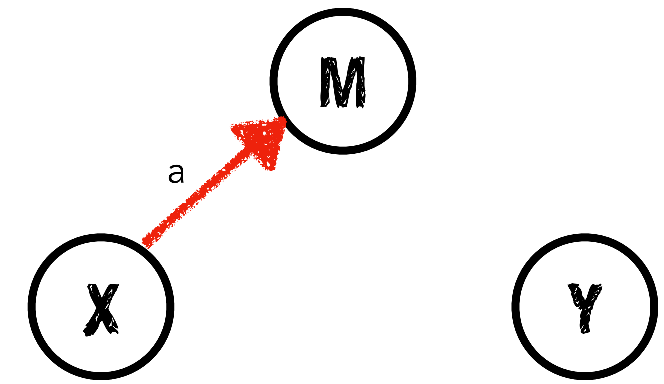

```r
# fit the model
fit.y_x = lm(formula = y ~ 1 + x,
             data = df.mediation)

# summarize the results
fit.y_x %>% summary()
```

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

Residual standard error: 5.16 on 98 degrees of freedom
Multiple R-squared:  0.04237,	Adjusted R-squared:  0.0326 
F-statistic: 4.336 on 1 and 98 DF,  p-value: 0.03993
```

significant relationship

Is there a relationship between **X** and **M**?

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



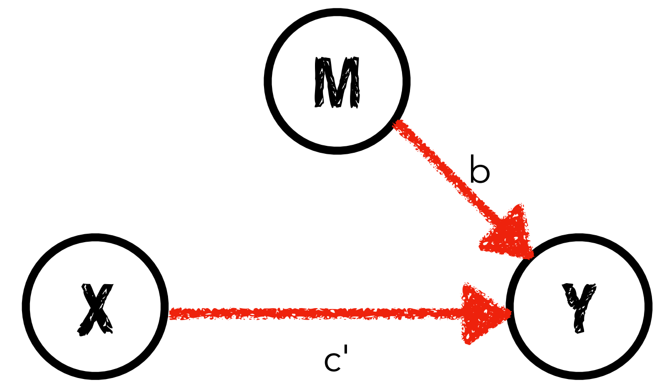

```r
# fit the model
fit.m_x = lm(formula = m ~ 1 + x,
             data = df.mediation)

# summarize the results
fit.m_x %>% summary()
```

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

Residuals:
    Min      1Q  Median      3Q     Max 
-9.5367 -3.4175 -0.4375  2.9032 16.4520 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.04494   13.41692   0.451    0.653    
x            0.66252    0.07634   8.678 8.87e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.854 on 98 degrees of freedom
Multiple R-squared:  0.4346,	Adjusted R-squared:  0.4288 
F-statistic: 75.31 on 1 and 98 DF,  p-value: 8.872e-14
```

significant relationship

Is there a relationship between **X** and **Y,** controlling for **M**?

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

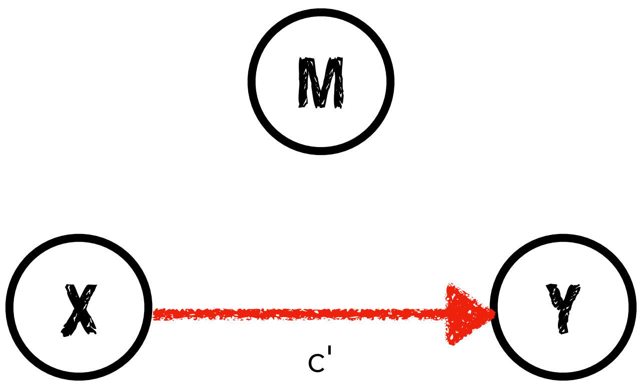



```r
# fit the model
fit.y_mx = lm(formula = m ~ 1 + m + x,
              data = df.mediation)

# summarize the results
fit.y_mx %>% summary()
```

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

Residual standard error: 4.756 on 97 degrees of freedom
Multiple R-squared:  0.1946,	Adjusted R-squared:  0.1779 
F-statistic: 11.72 on 2 and 97 DF,  p-value: 2.771e-05
```

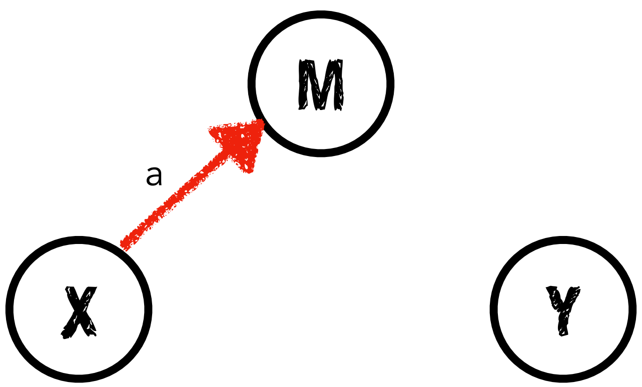
not significant

3 Step procedure



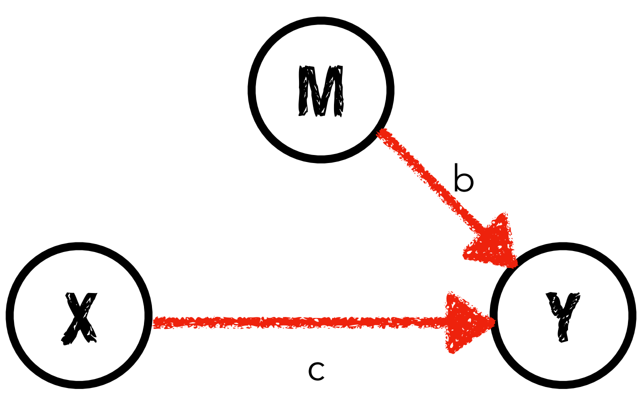
Relationship between **X** and **Y**?

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



Relationship between **X** and **M**?

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



Relationship between **X** and **Y**, controlling for **M**?

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

significant change?

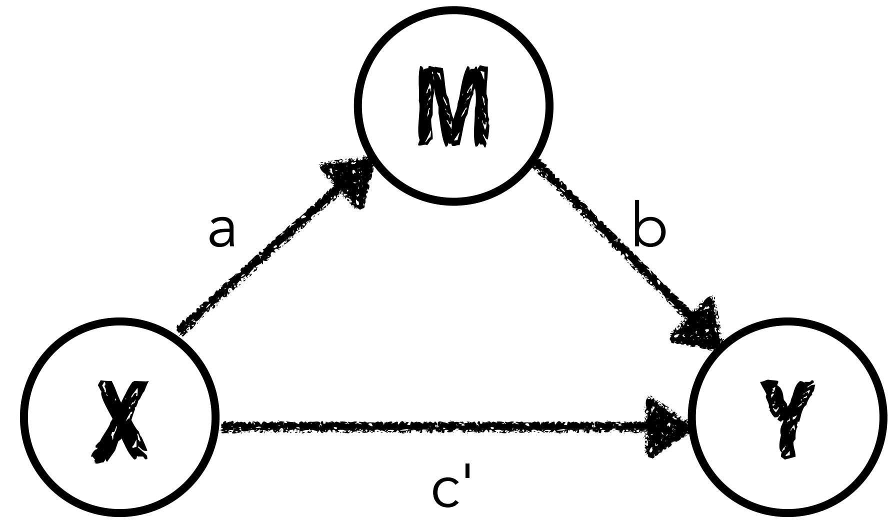
Mantra:

a change in significance is not always a significant change

Is the mediation significant?

1. **Sobel test**
    - assumes normally distributed data
    - has low power
2. **Bootstrapping**
    - no assumption about how the data is distributed
    - has higher power
3. **Bayesian mediation analysis**

1. Sobel test

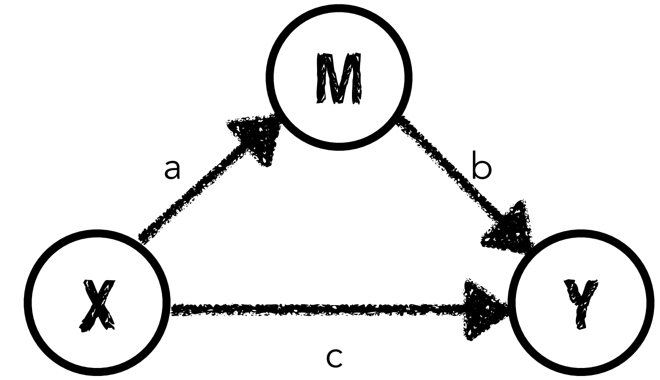



```r
library("multilevel")

# run the sobel test
fit.sobel = sobel(pred = df.mediation$x,
                  med = df.mediation$m,
                  out = df.mediation$y)

# calculate the p-value
(1 - pnorm(fit.sobel$z.value))*2
```

"It is becoming increasingly more difficult to publish tests of mediation based purely on the Baron and Kenny method or tests that make distributional assumptions such as the Sobel test."

A Wikipedia person

2. Bootstrapping



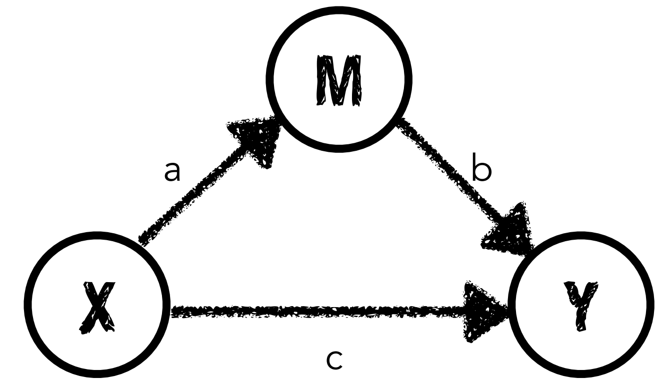

```r
library("mediation")

# 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()
```

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

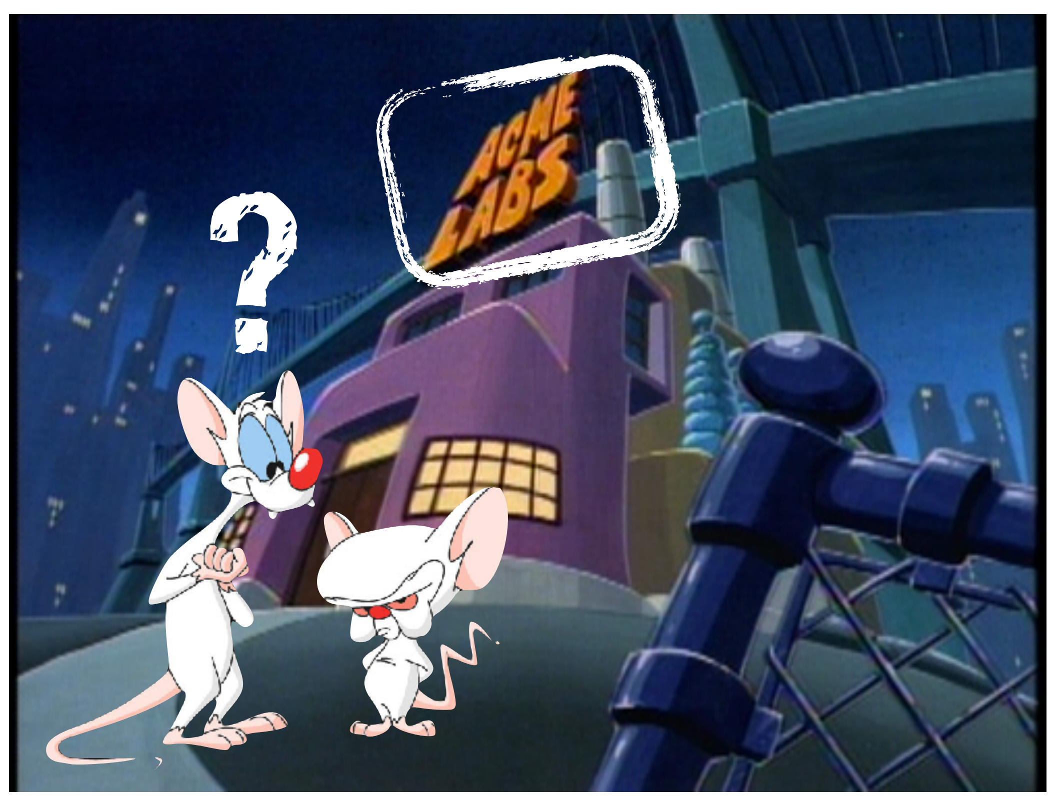
$$\hat{y} = b_0 + b_1 \cdot m + b_2 \cdot x$$

```
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
```

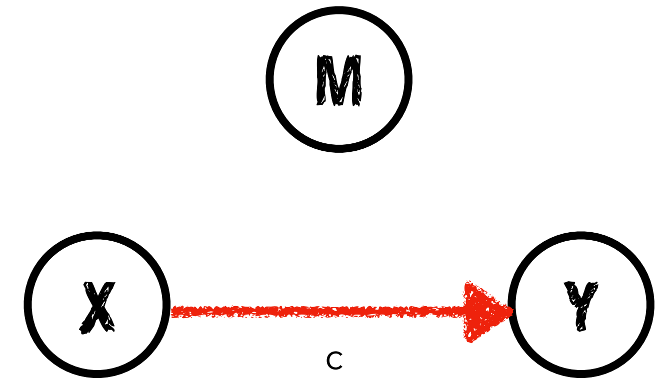
2. Bootstrapping




```
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

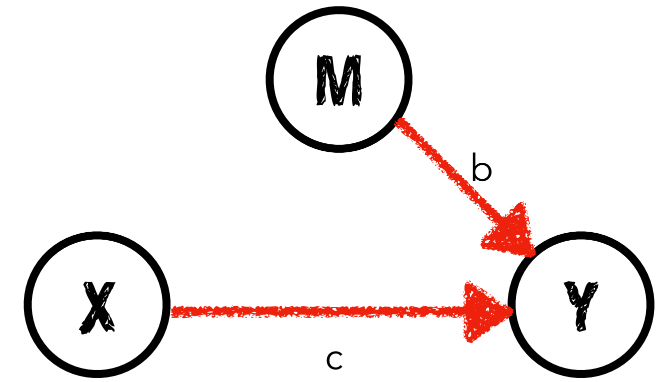



```
Causal Mediation Analysis

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

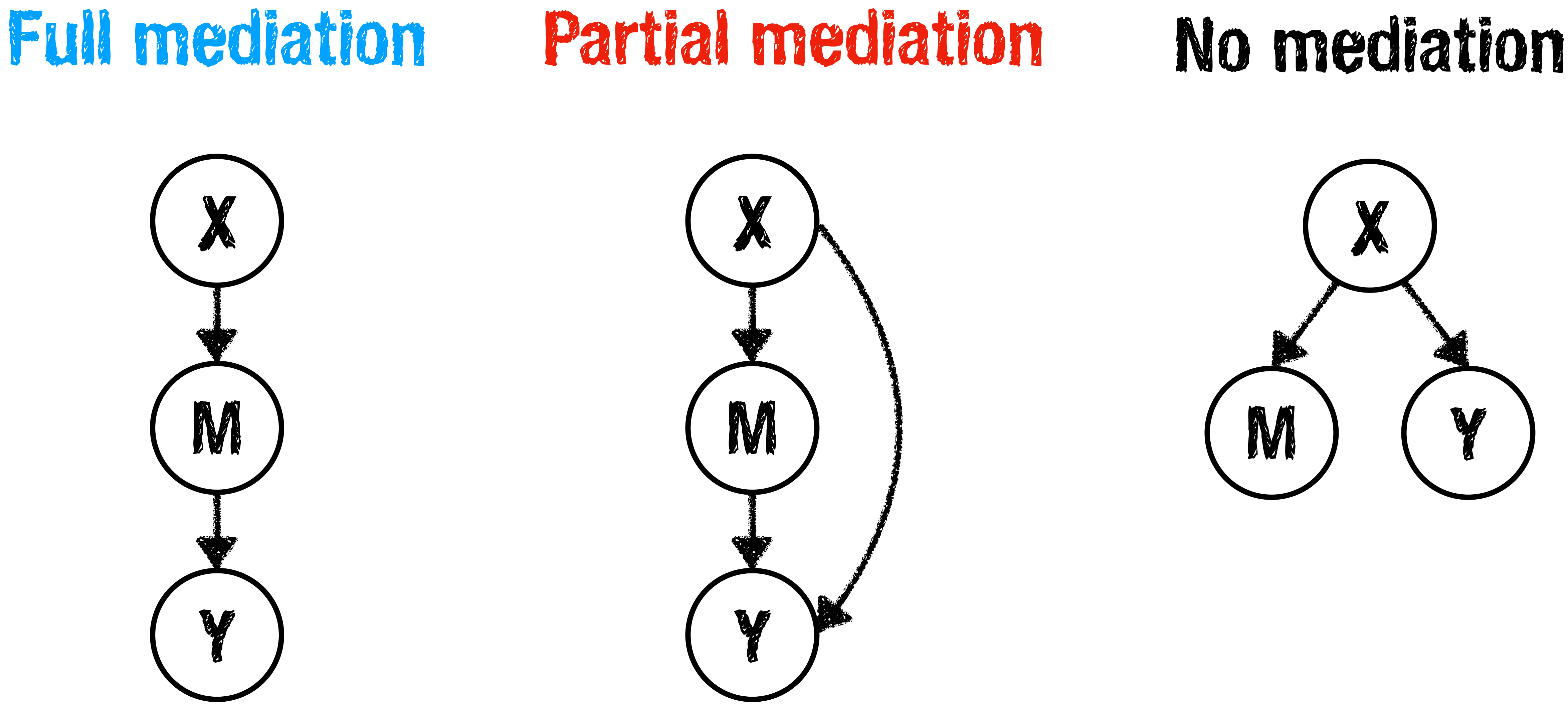
```
Causal Mediation Analysis

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

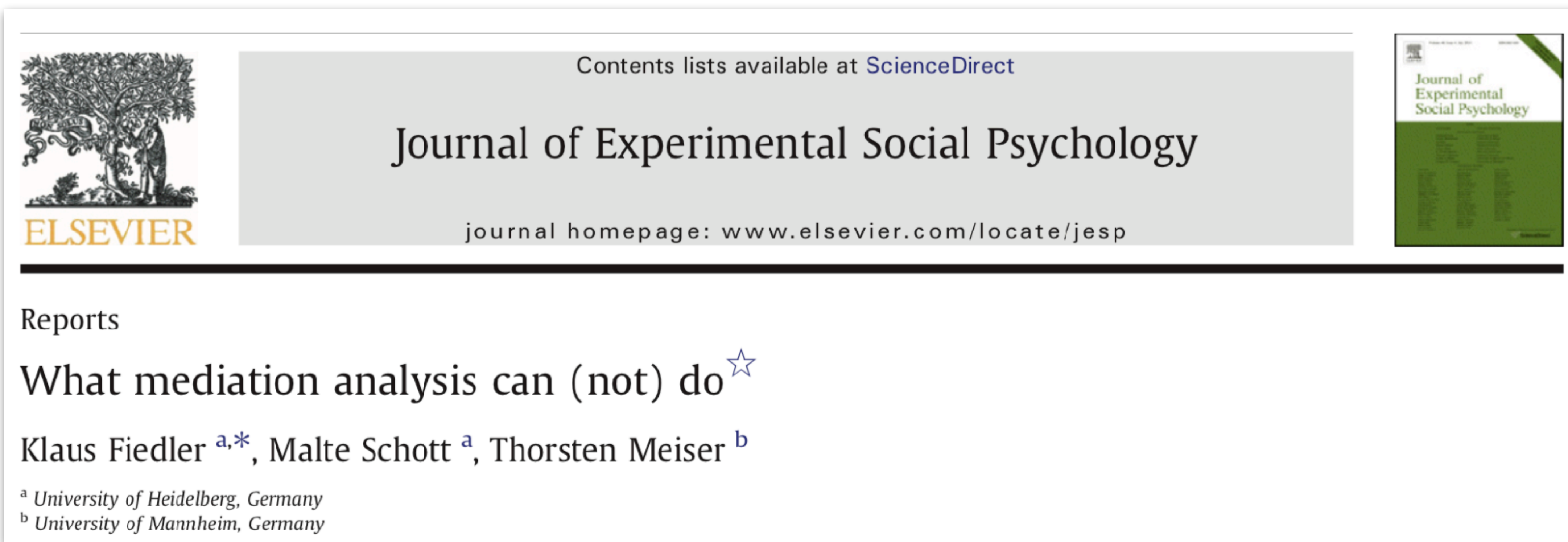


Full mediation: When the effect of **X** on **Y** completely disappears, **M** fully mediates between **X** and **Y**.

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

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

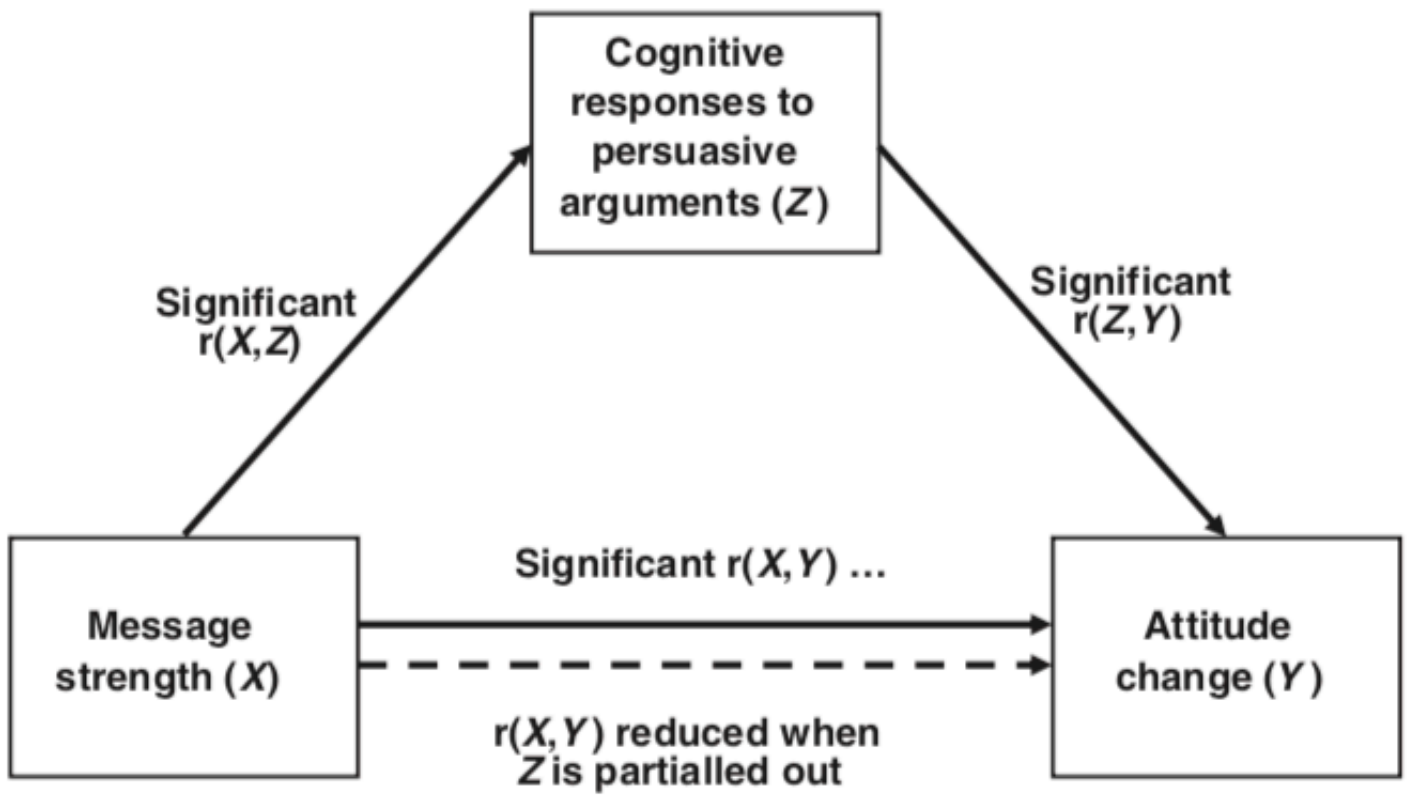


Fig. 1. Illustration of a mediation analysis in the context of persuasion research: The impact of message strength (X) on attitude change (Y) is supposed to be mediated by the recipients' cognitive responses (Z).

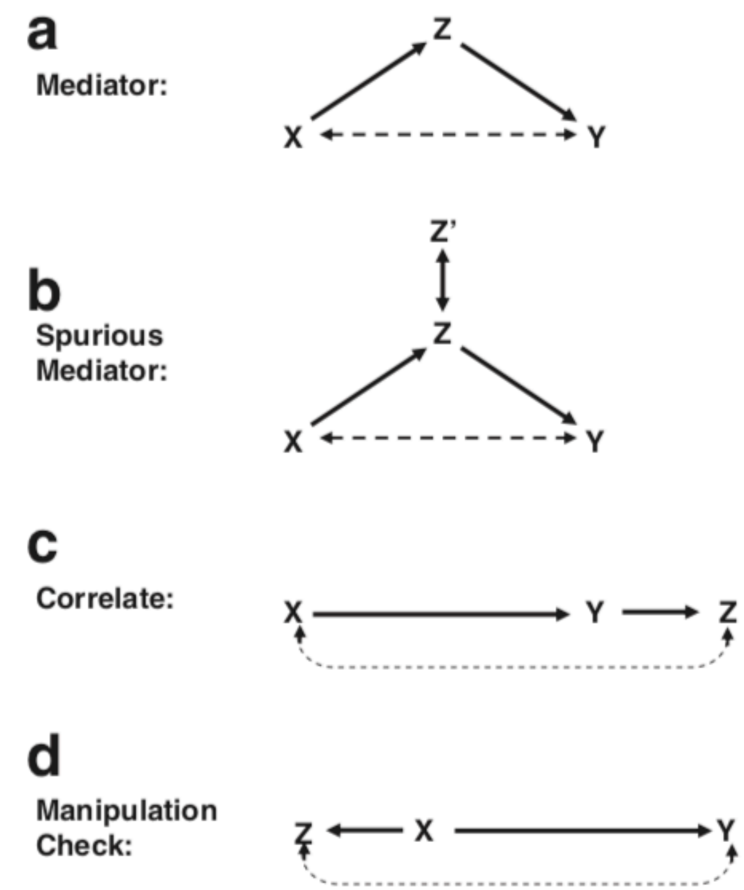


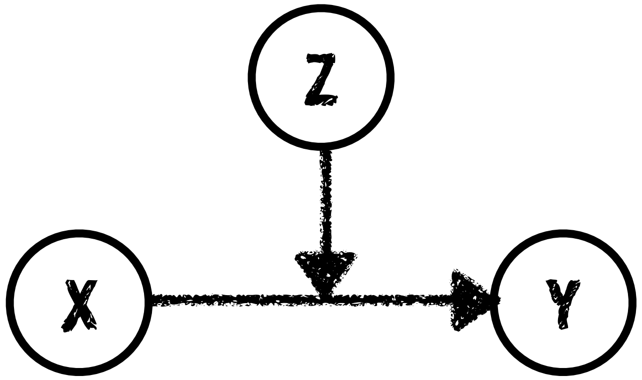
Fig. 2. Different causal models for the interpretation of the same intercorrelations between three variables, X, Y, and Z.

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

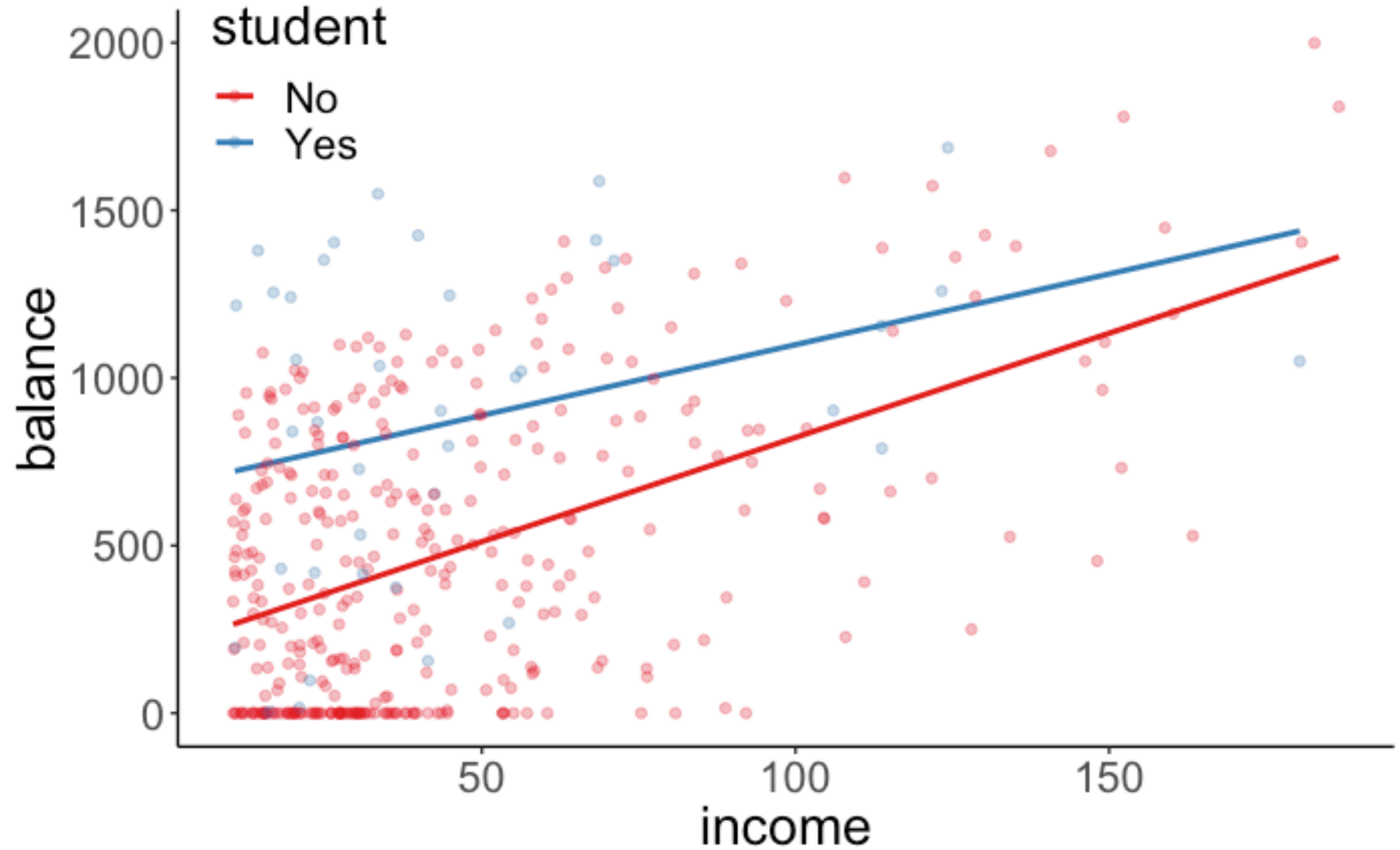


Moderation means that the effect of a predictor depends on the value of another.

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.



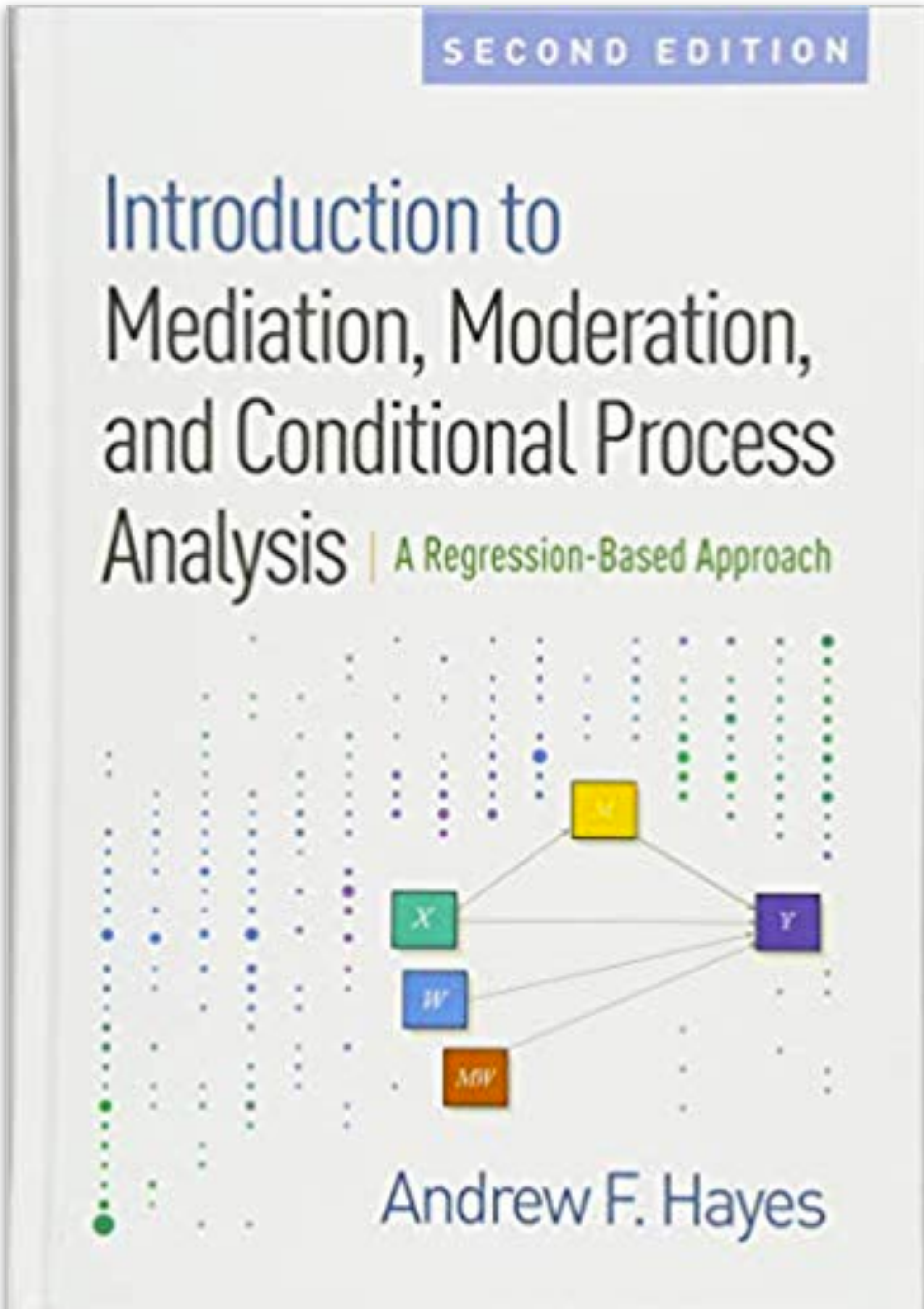
$$\widehat{\text{balance}}_i = 200.62 + 6.22 \cdot \text{income}_i + 476.68 \cdot \text{student}_i - 2.00 \cdot (\text{income}_i \times \text{student}_i)$$

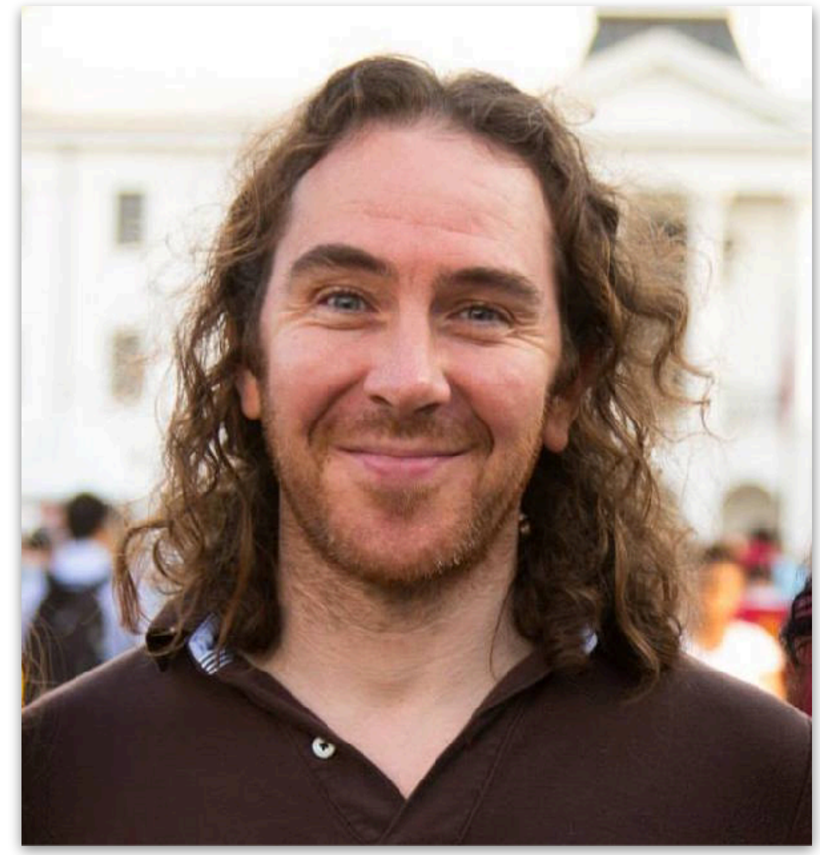
if student = "No" $\widehat{\text{balance}}_i = 200.62 + 6.22 \cdot \text{income}_i$

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





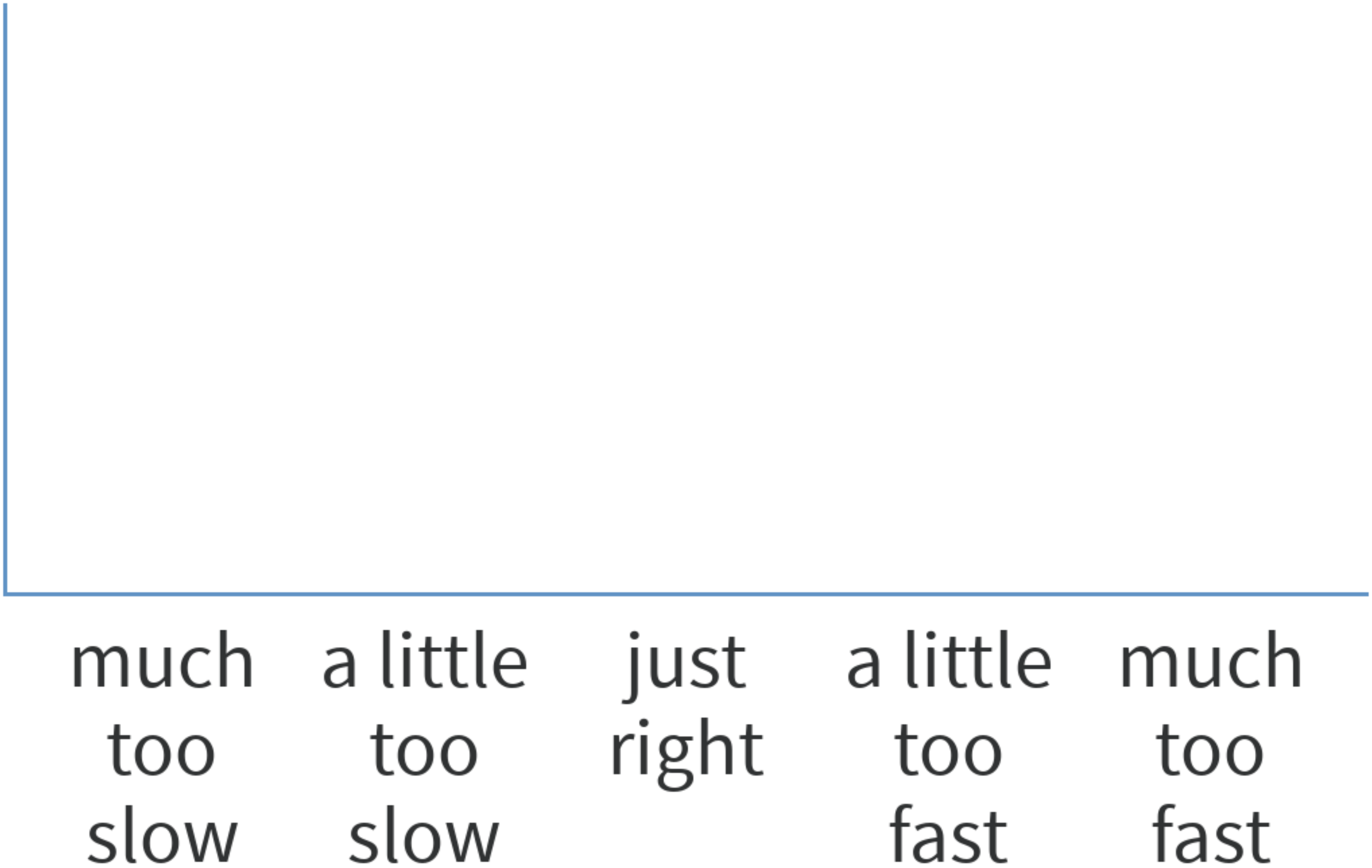
Recoded with `brms` by Solomon Kurz here: https://bookdown.org/connect/#/apps/1523/access

Plan for today

- Controlling for variables
- Mediation
- Moderation

Feedback

How was the pace of today's class?



How happy were you with today's class overall?

Start the presentation to see live content. Still no live content? Install the app or get help at **PollEv.com/app**

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

Thank you!