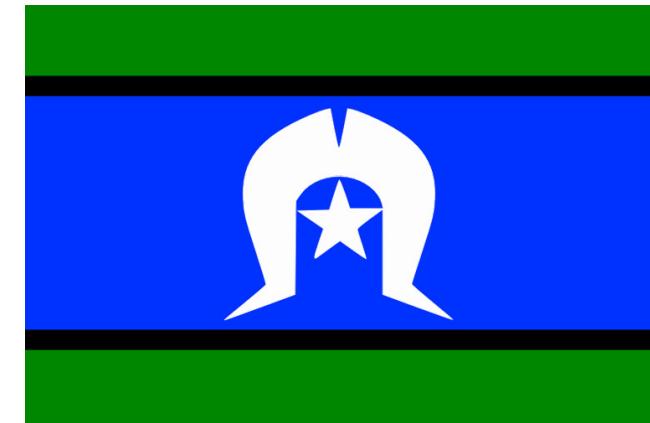
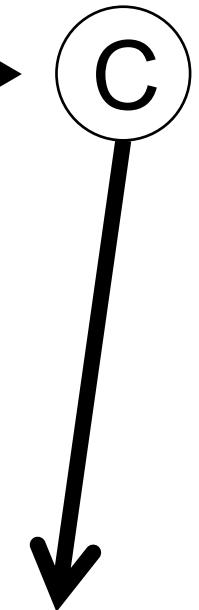


Kevin Bairos-Novak
CodeR-TSV Presentation – 11am Oct 25th, 2021
(heavily adapted from videos by Richard McElreath)



Acknowledgement of Country
Wulgurukaba and Bindal Peoples' Land

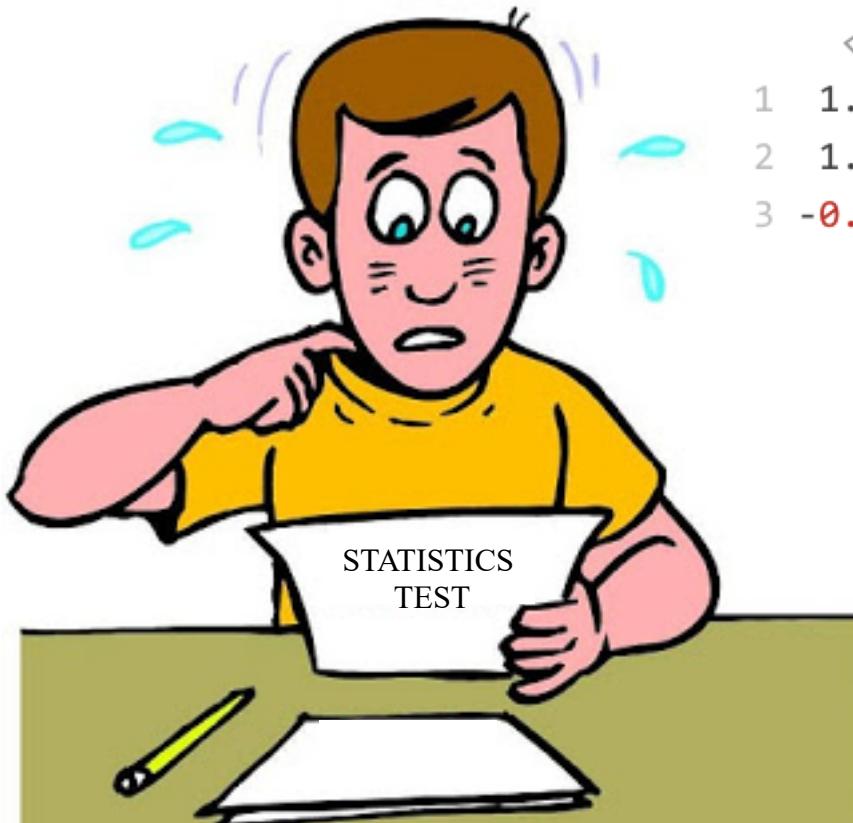


Chapter 1: The first step is acceptance.

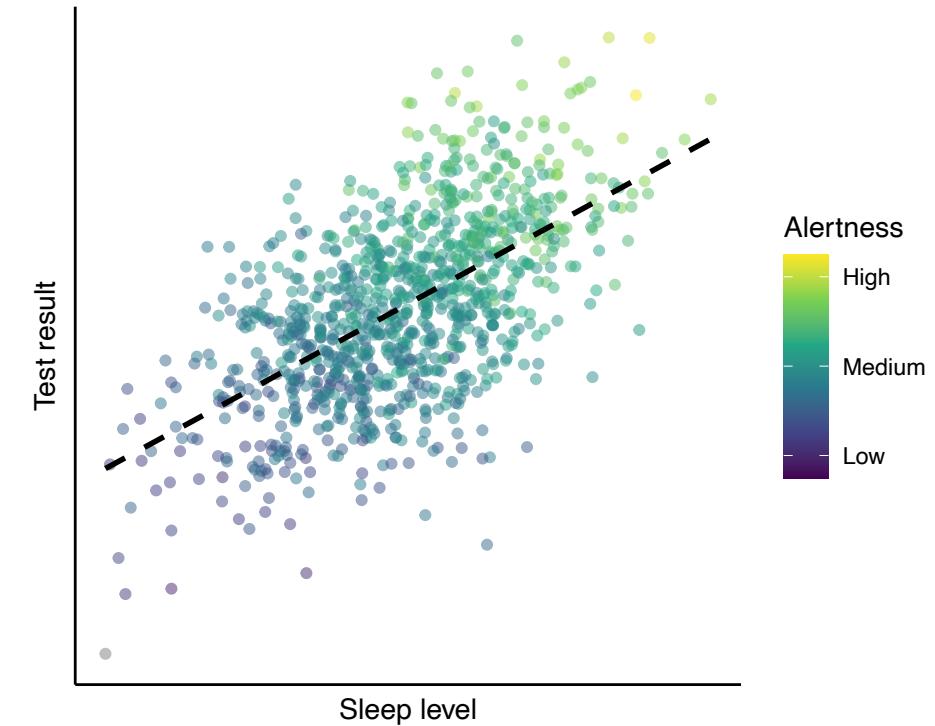


Does sleep level affect student test results?

No, alertness does though!



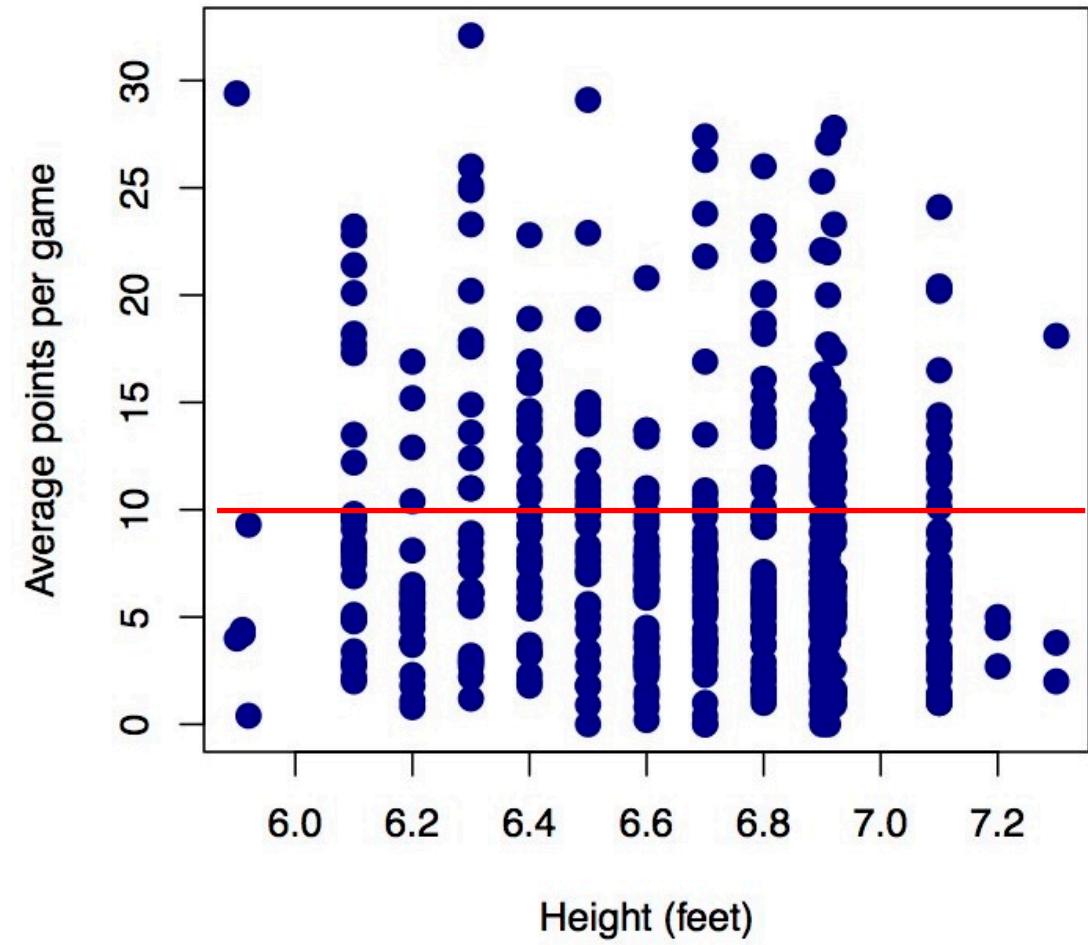
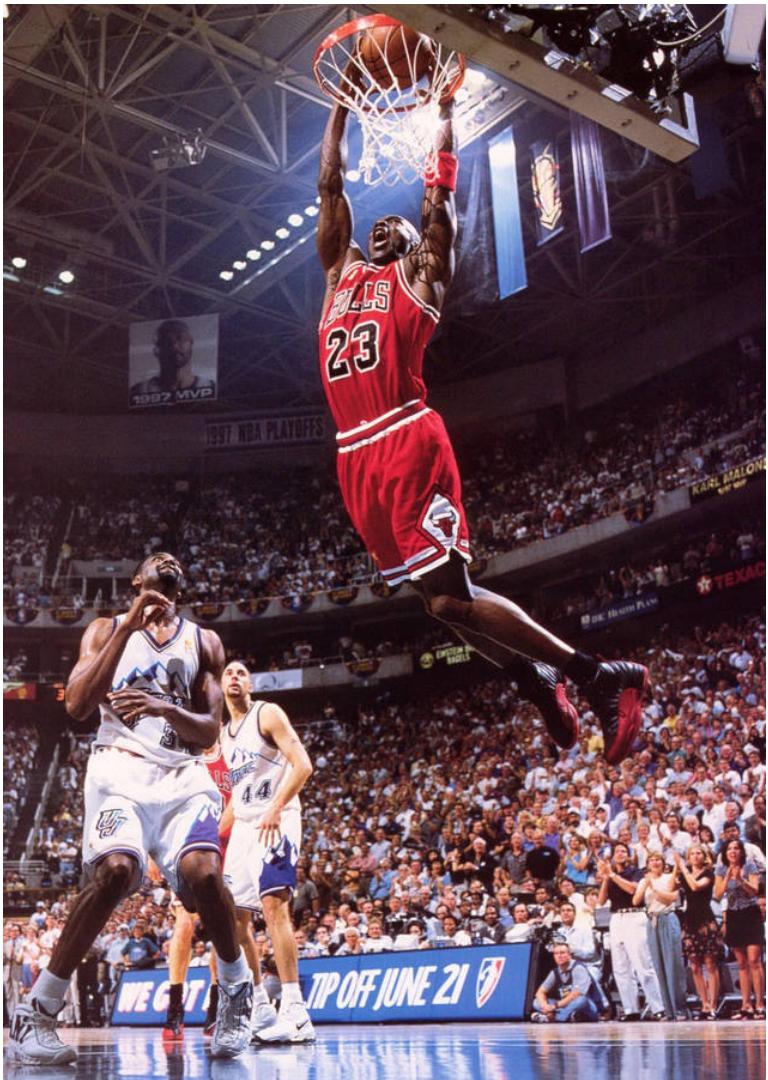
```
# A tibble: 6 x 3
  sleep  alertness test_result
  <dbl>    <dbl>      <dbl>
1 1.13     0.249     0.988
2 1.11    -0.810    -0.424
3 -0.871    0.749     2.05
```



Coefficients:

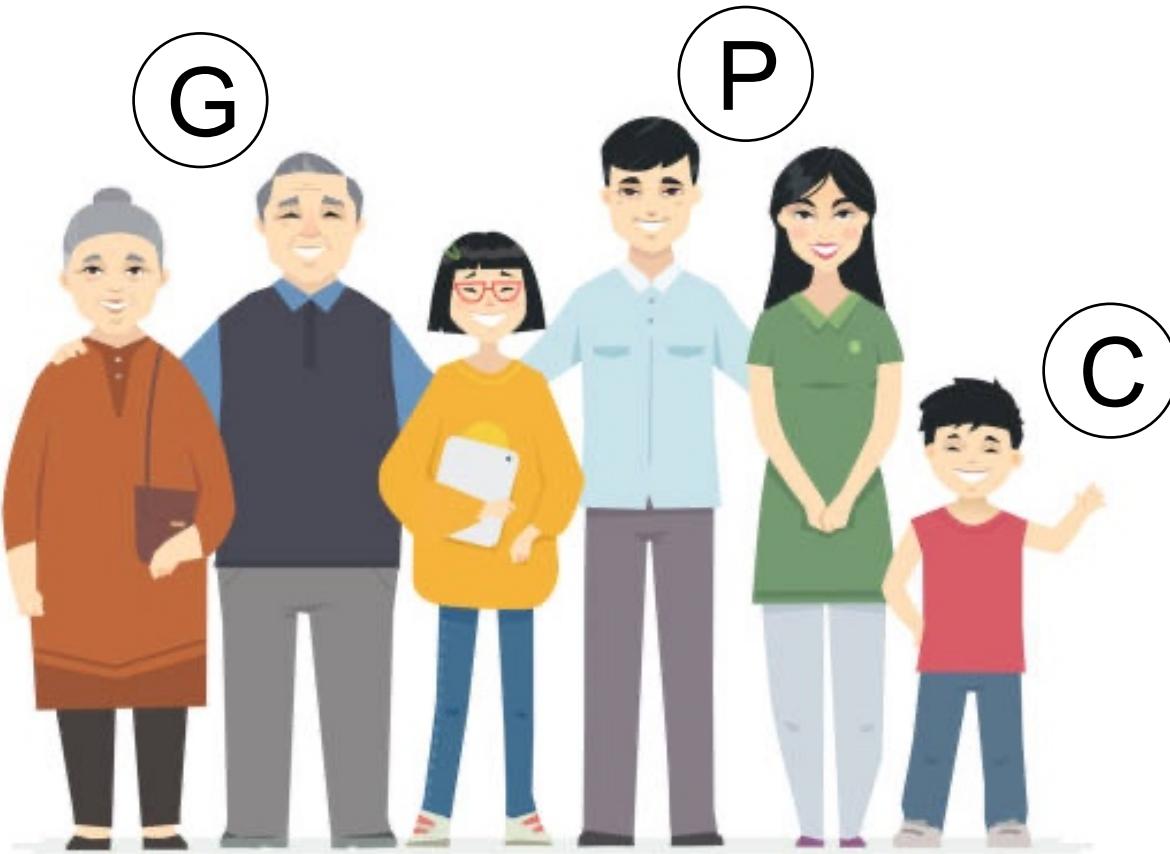
| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|------------|
| (Intercept) | 0.016811 | 0.032880 | 0.511 | 0.609 |
| sleep | 0.004877 | 0.045431 | 0.107 | 0.915 |
| alertness | 1.010407 | 0.031905 | 31.670 | <2e-16 *** |
| --- | | | | |

Does basketball player height affect scoring?

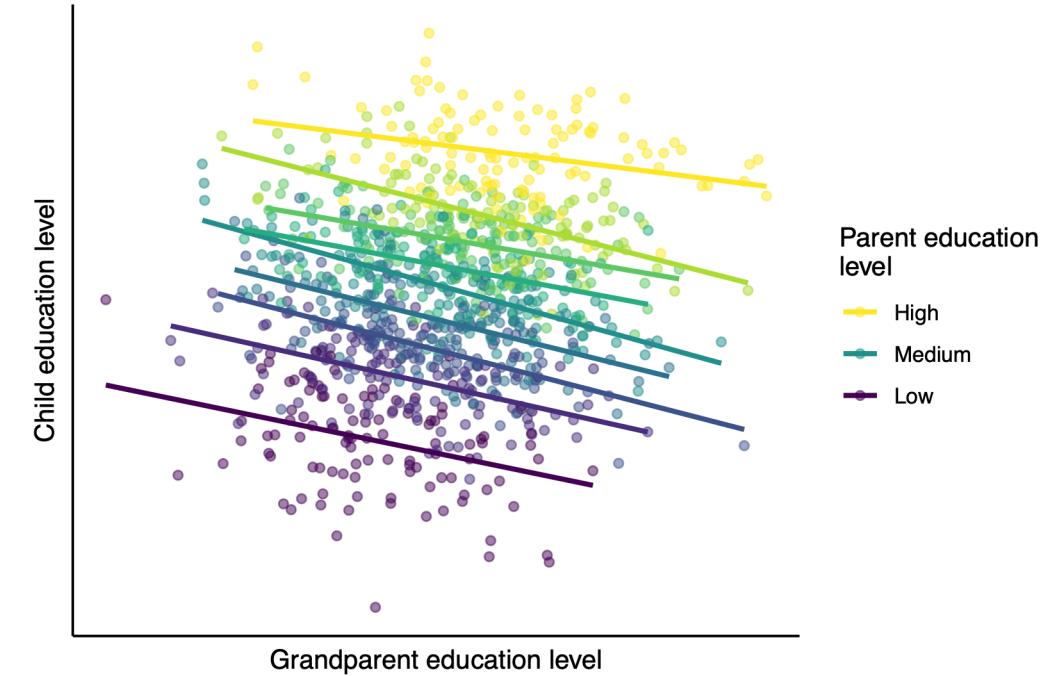


No relationship between height
and scoring ability!

Does your grandparent's education level affect your own?



Grandparents' education has a
NEGATIVE effect on grandchild's!



Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 0.09264 | 0.19290 | 0.480 | 0.631 |
| G | 0.79952 | 0.19050 | 4.197 | 2.95e-05 *** |

Coefficients:

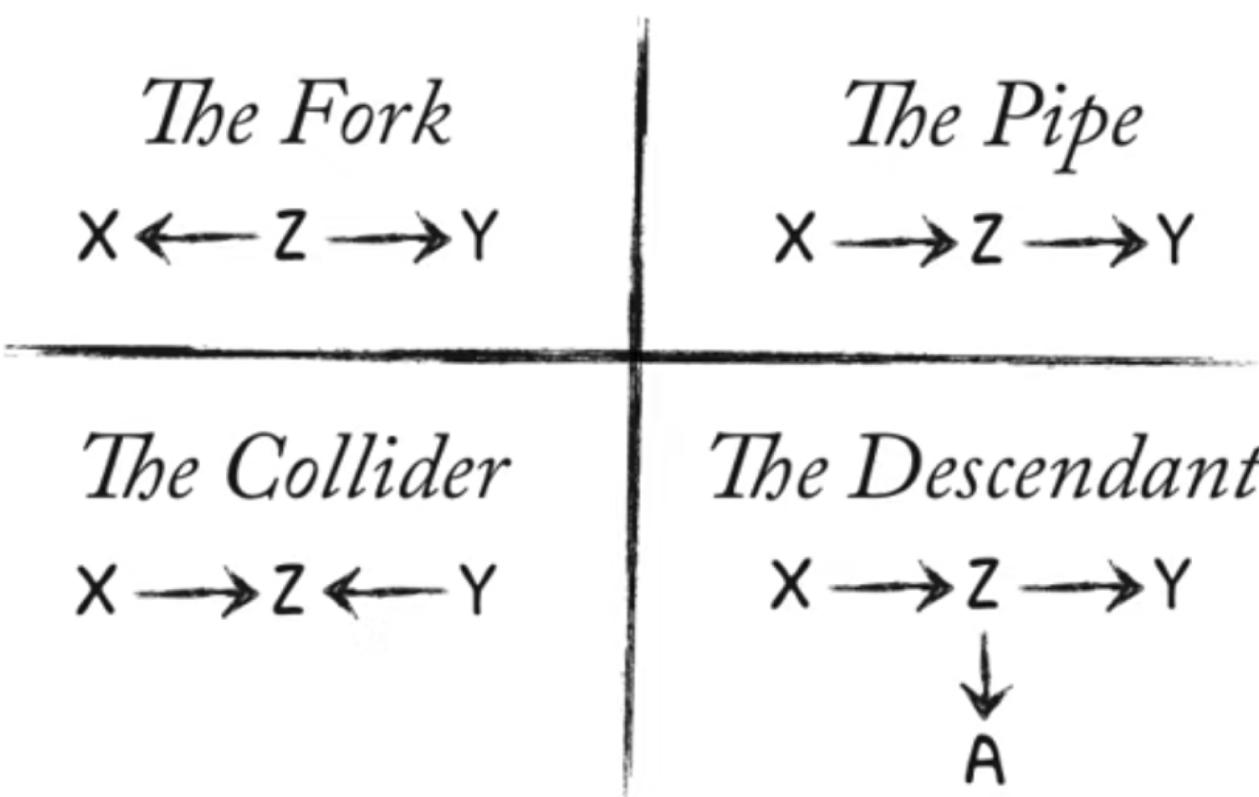
| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|-----------|------------|---------|------------|
| (Intercept) | -0.020116 | 0.055275 | -0.364 | 0.716 |
| G | -1.235624 | 0.053557 | -23.071 | <2e-16 *** |
| P | 2.209555 | 0.022349 | 98.864 | <2e-16 *** |
| G:P | 0.003598 | 0.018804 | 0.191 | 0.848 |

We have a problem. The first step is acceptance.

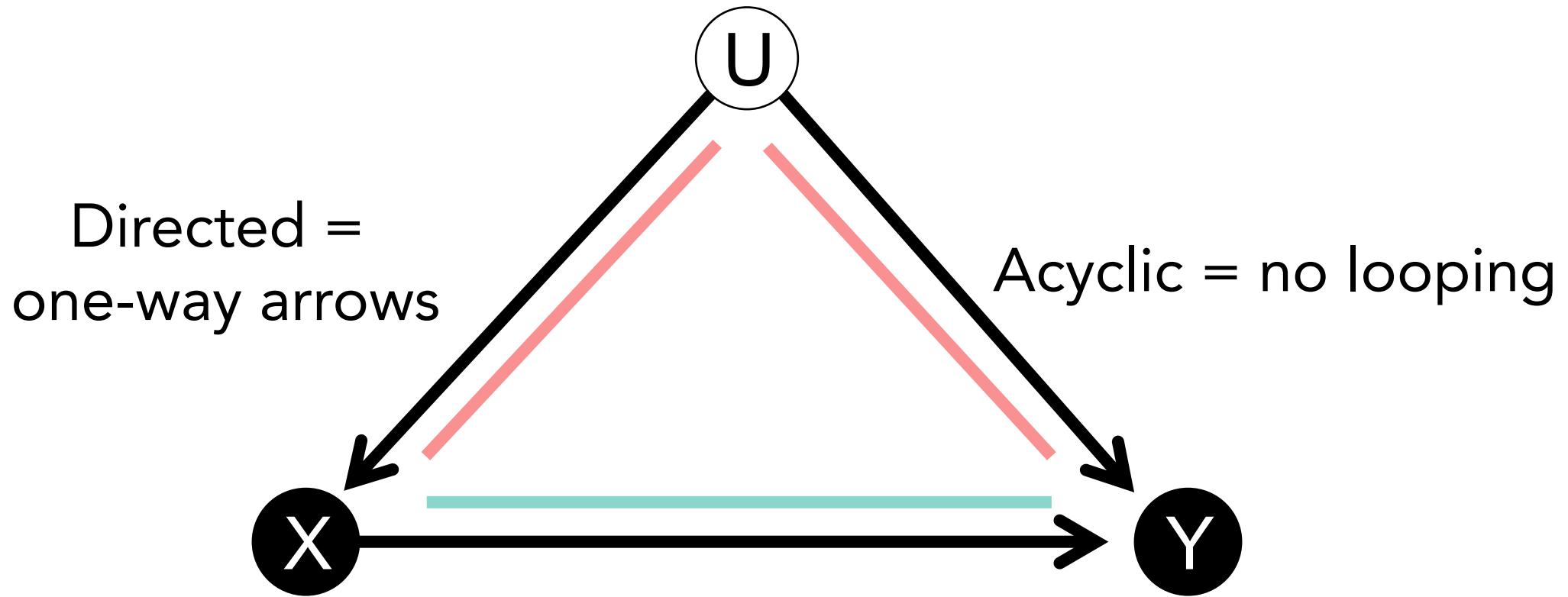


Chapter 2: Embracing alchemy.

The Four Elemental Confounds



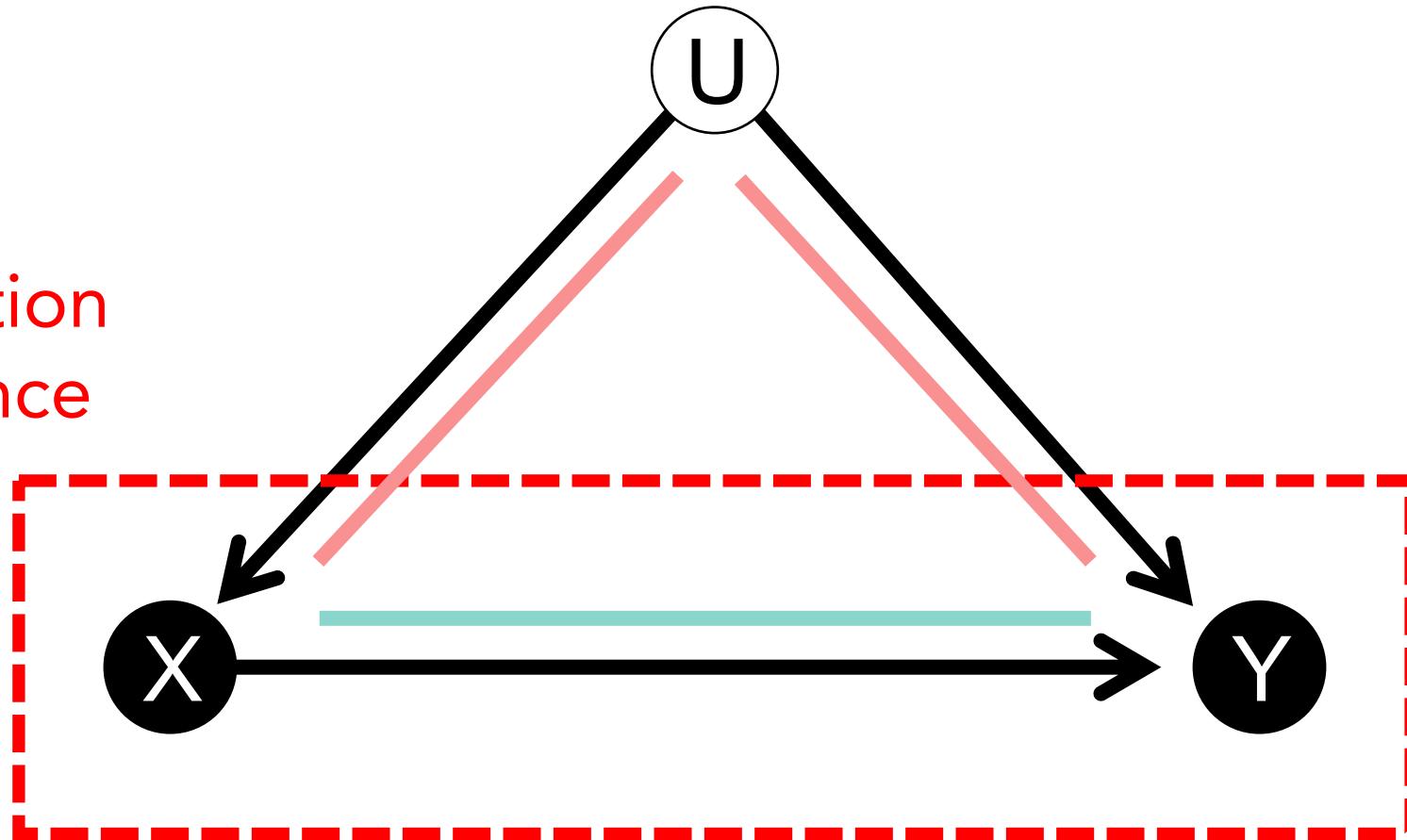
Confounds are everywhere.
DAGs (Directed Acyclical Graphs) can help.



Graph = a late 19th century abbreviation of "graphic formula"

Good experimental design can help eliminate unknown confounds.

Randomization
Independence
Replication
Blocking
Controls



When we can't manipulate, we need causal alchemy!

The Fork

$$X \leftarrow Z \rightarrow Y$$

Open unless you
condition on Z

The Pipe

$$X \rightarrow Z \rightarrow Y$$

Open unless you
condition on Z

The Collider

$$X \rightarrow Z \leftarrow Y$$

Closed until you
condition on Z

The Descendant

$$X \rightarrow Z \rightarrow Y$$

↓
A

Conditioning on A is
like conditioning on Z

When we can't manipulate, we need causal alchemy!

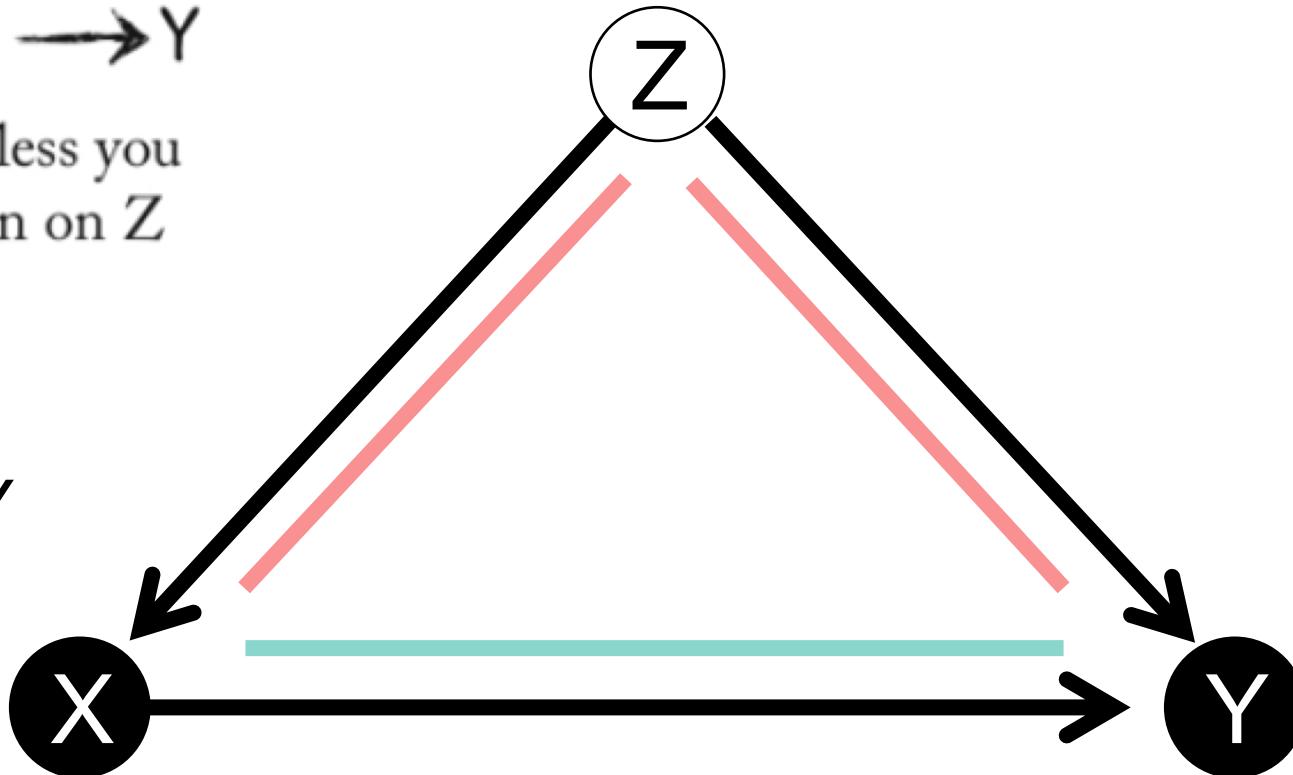
The Fork

$$X \leftarrow Z \rightarrow Y$$

Open unless you
condition on Z

Two paths:

- Path A: $X \rightarrow Y$
- Path B: $X \leftarrow Z \rightarrow Y$



When we can't manipulate, we need causal alchemy!

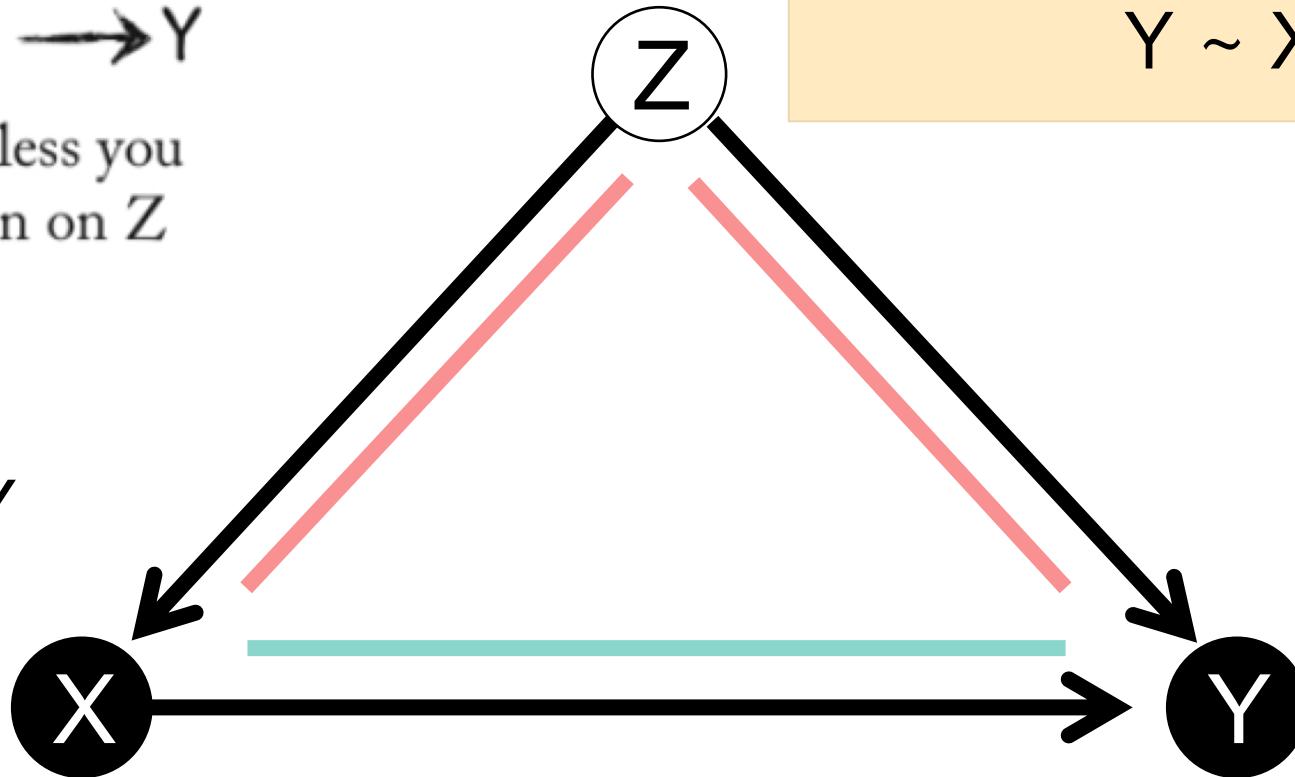
The Fork

X ← Z → Y

Open unless you
condition on Z

Two paths:

- Path A: X → Y
 - Path B: X ← Z → Y



Not ideal model: $Y \sim X$

There is a backdoor path through Z that must be closed!

When we can't manipulate, we need causal alchemy!

The Fork

$$X \leftarrow Z \rightarrow Y$$

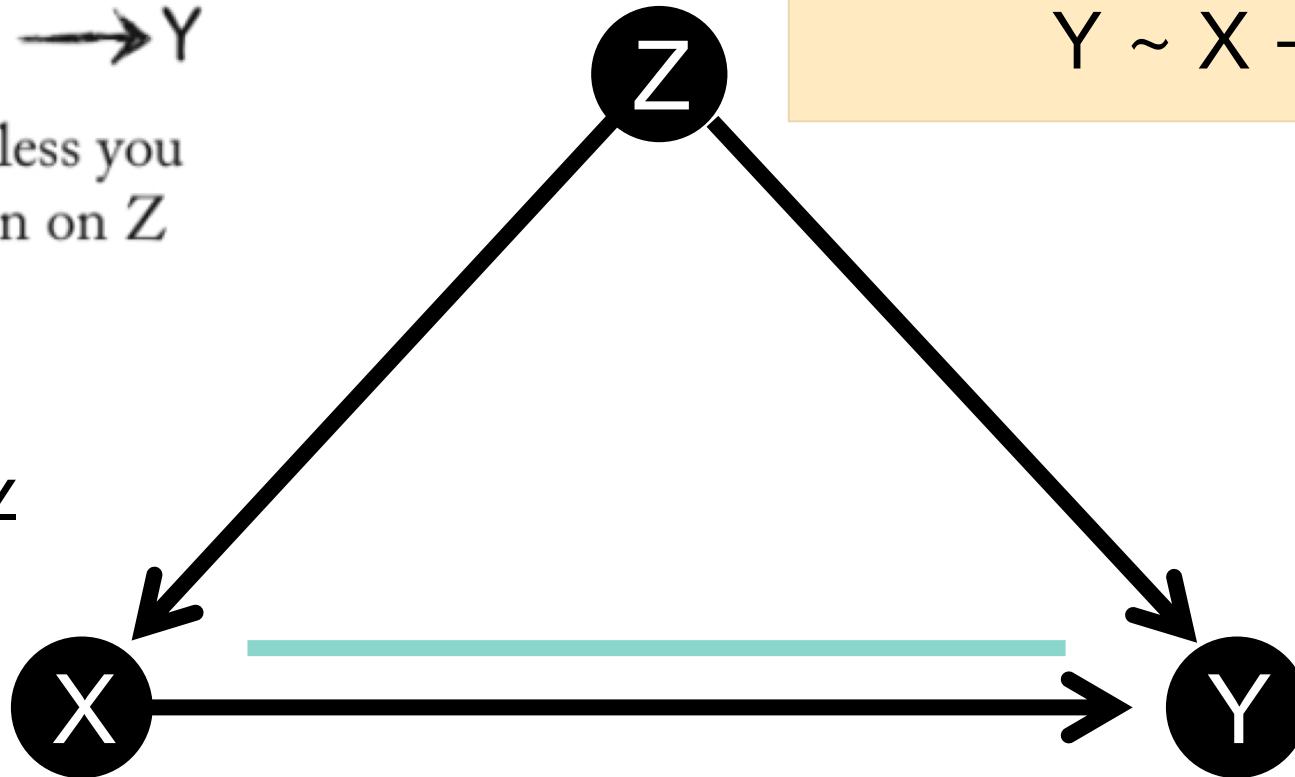
Open unless you condition on Z

Two paths:

- Path A: $X \rightarrow Y$
- Path B: $X \leftarrow Z \rightarrow Y$

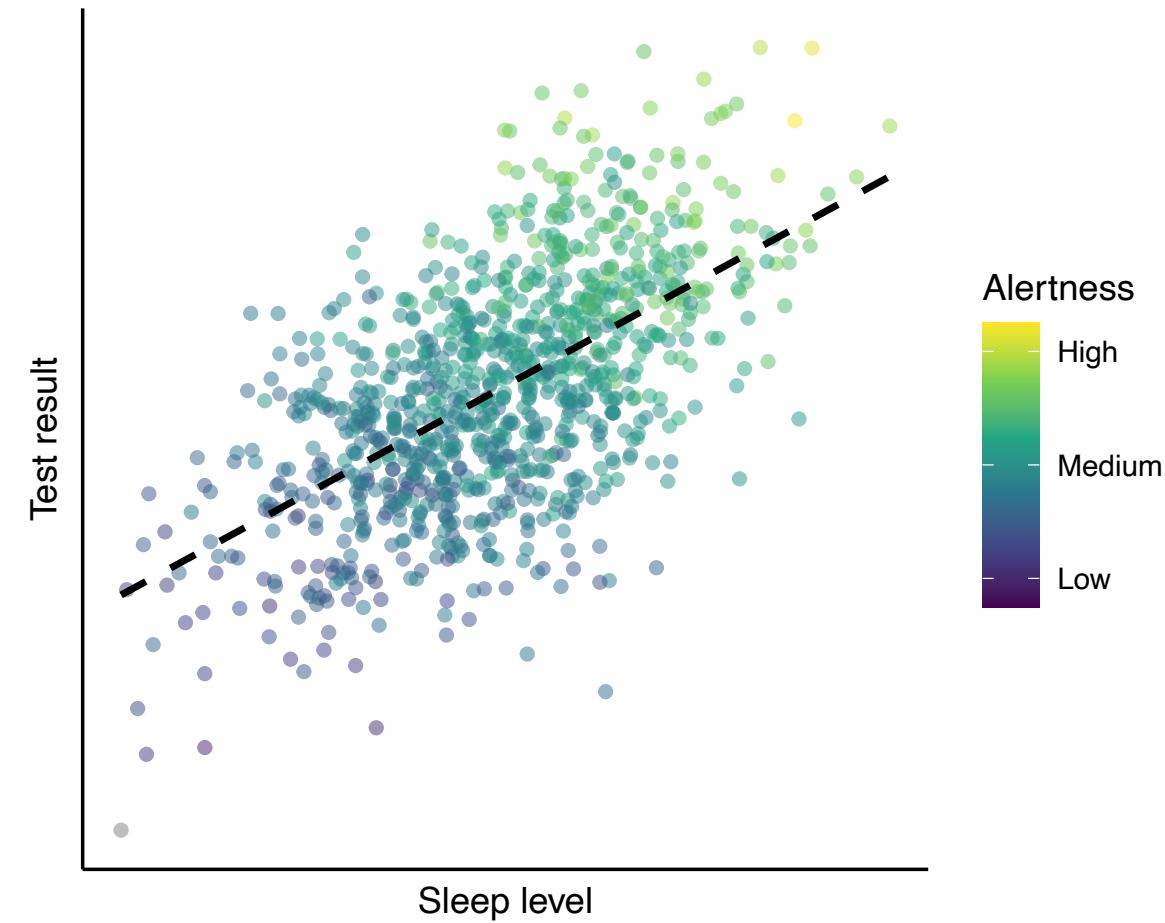
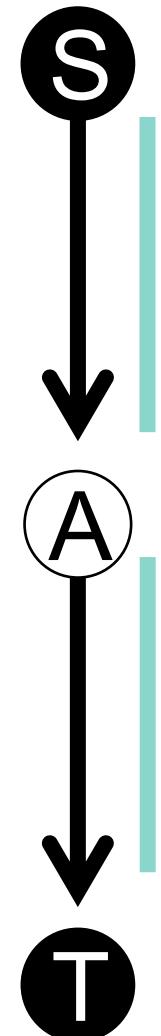
Ideal statistical model:

$$Y \sim X + Z$$



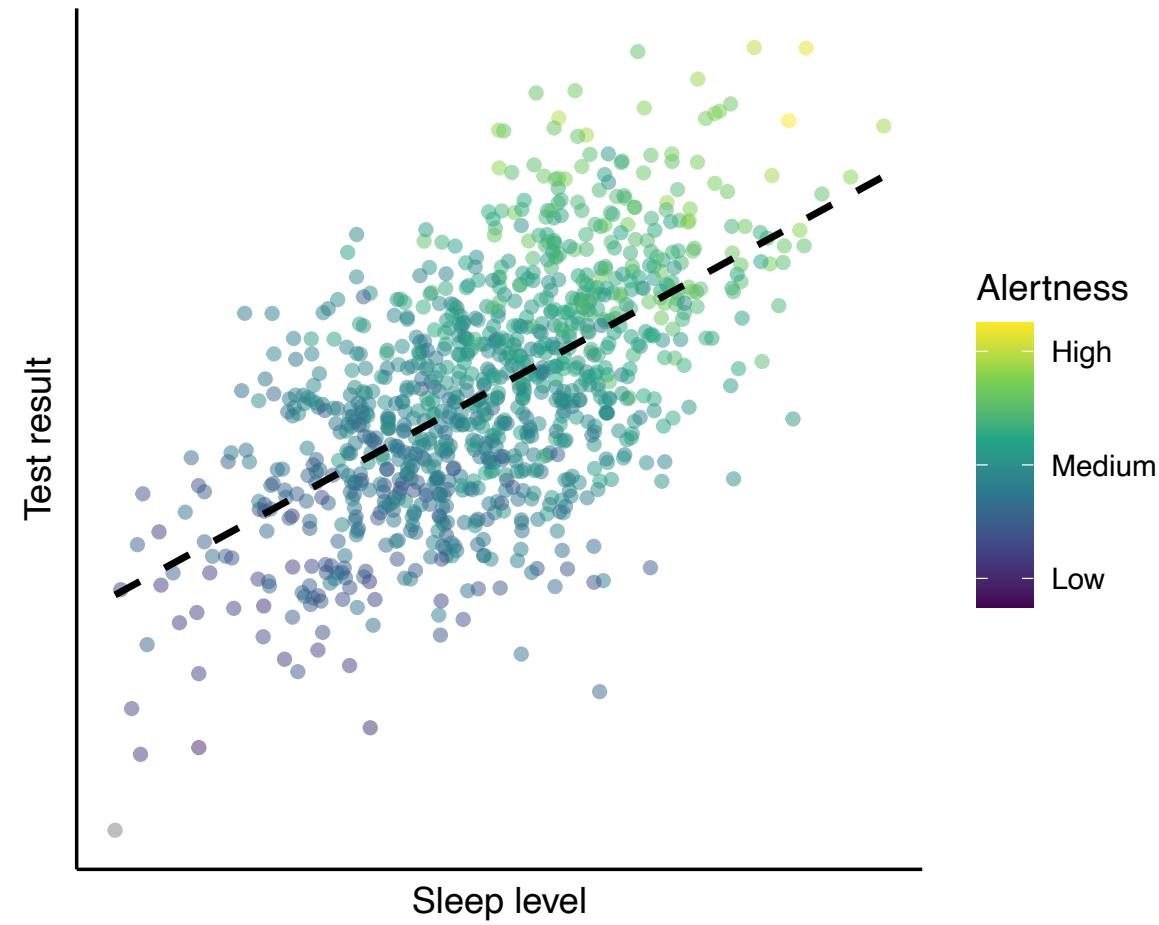
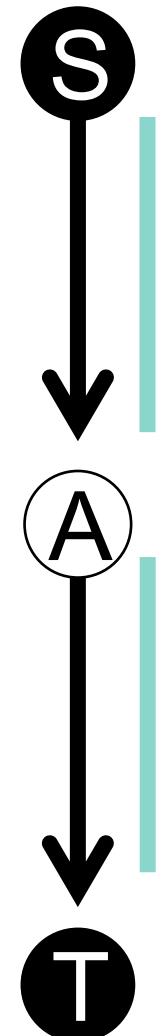
There is a backdoor path through Z that must be closed!

Does sleep level affect student test results?



One path:
• $S \rightarrow A \rightarrow T$

Does sleep level affect student test results?



One path:

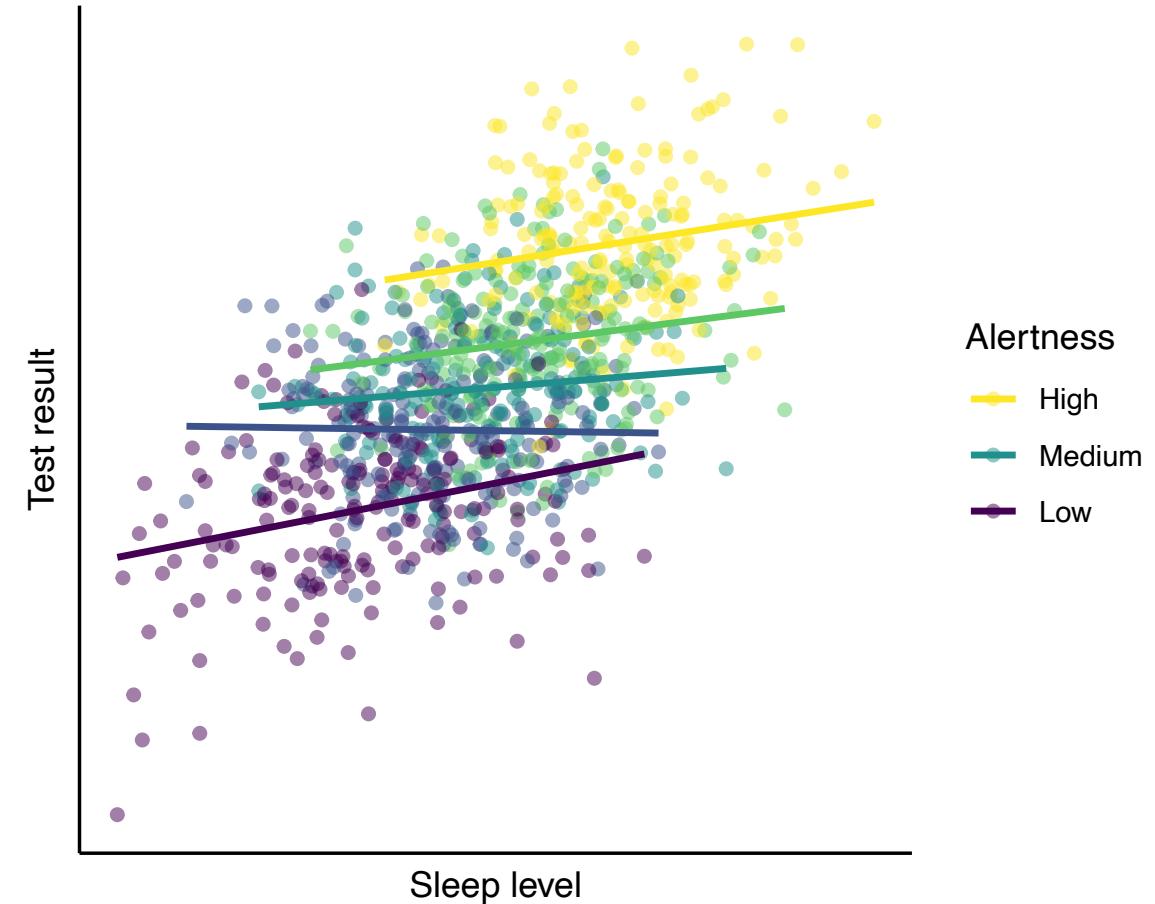
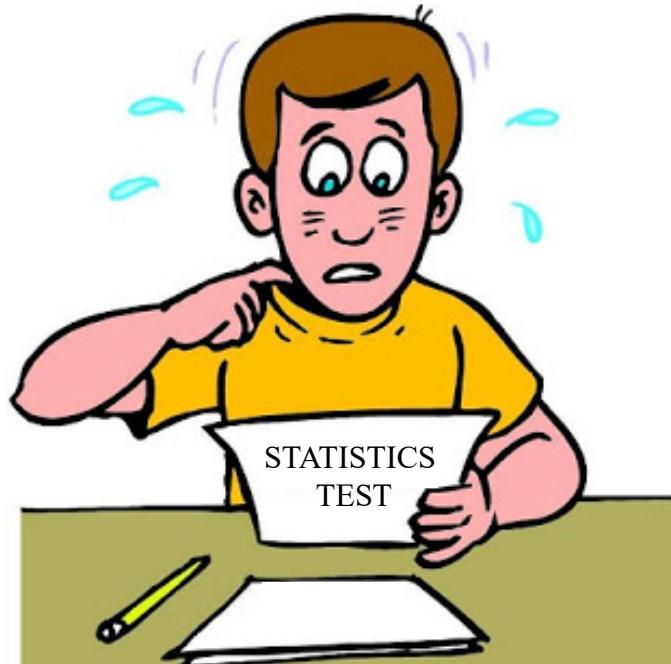
- $S \rightarrow A \rightarrow T$

Does sleep level affect student test results?

The Pipe

$$X \rightarrow Z \rightarrow Y$$

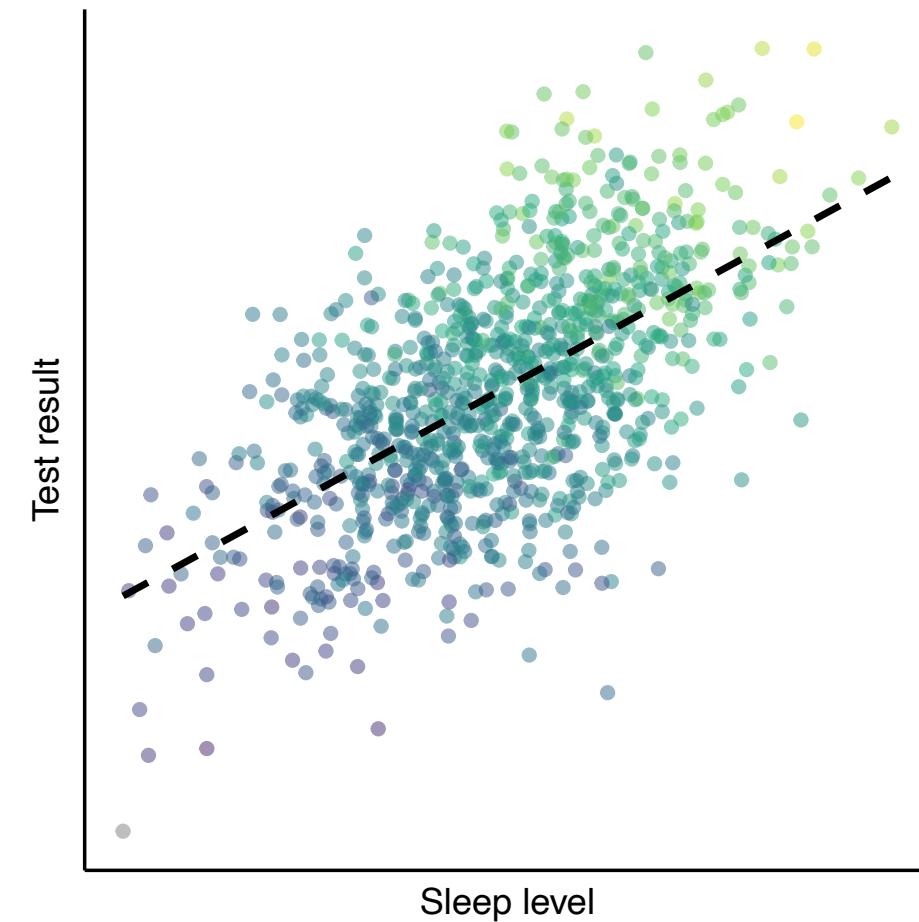
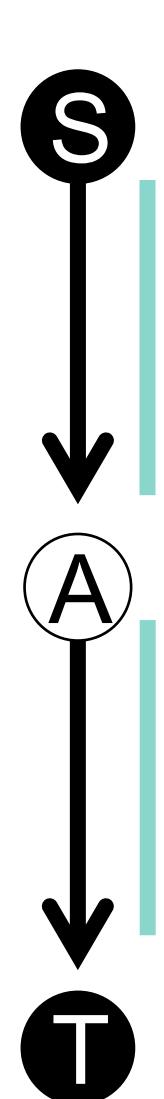
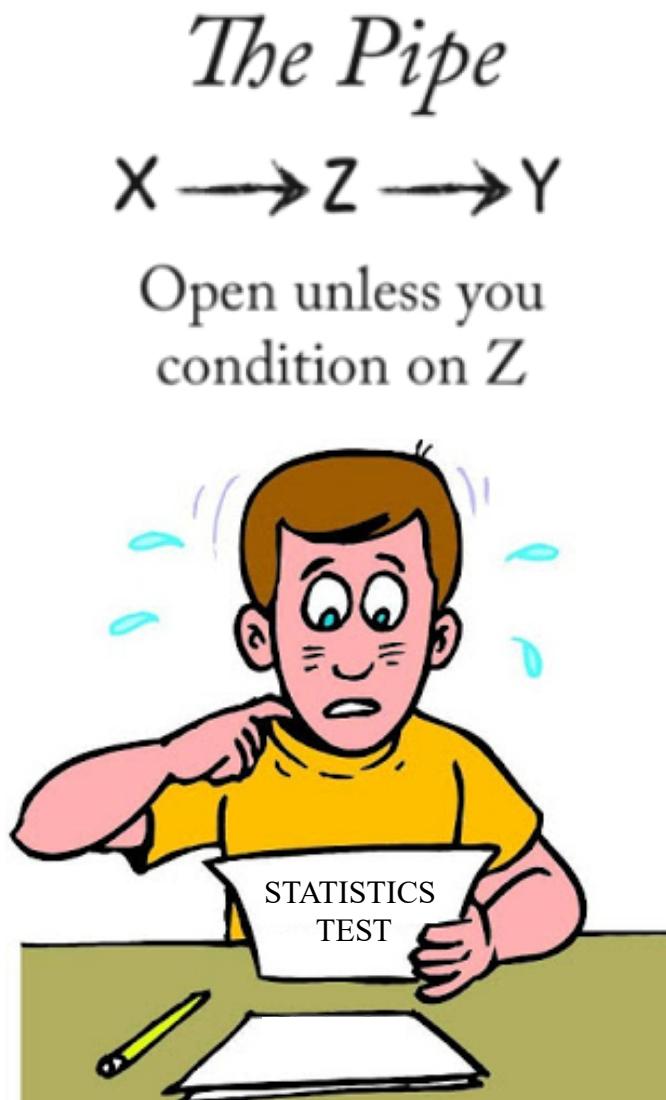
Open unless you
condition on Z



Not ideal model:
 $T \sim S + A$

| Coefficients: | | | | | |
|---------------|----------|------------|---------|------------|--|
| | Estimate | Std. Error | t value | Pr(> t) | |
| (Intercept) | 0.016811 | 0.032880 | 0.511 | 0.609 | |
| sleep | 0.004877 | 0.045431 | 0.107 | 0.915 | |
| alertness | 1.010407 | 0.031905 | 31.670 | <2e-16 *** | |
| | --- | | | | |

Does sleep level affect student test results? **YES!**



Ideal model:
 $T \sim S$

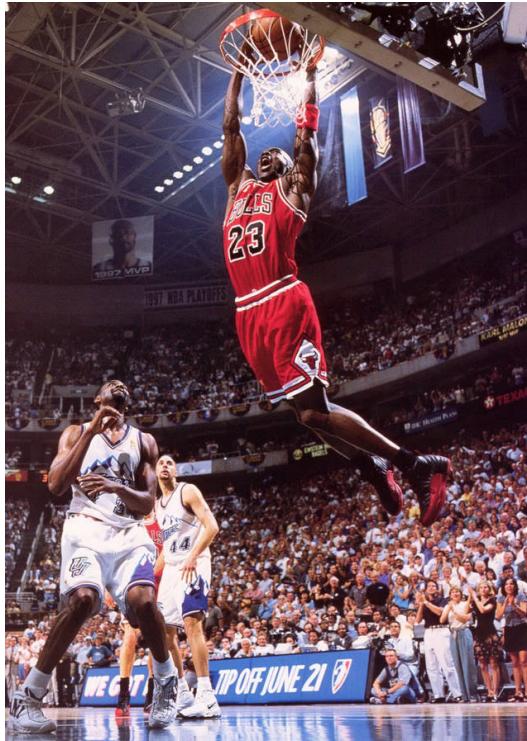
| | Coefficients: | | | | |
|-------------|---------------|------------|---------|------------|--|
| | Estimate | Std. Error | t value | Pr(> t) | |
| (Intercept) | 0.09989 | 0.04419 | 2.26 | 0.024 * | |
| sleep | 1.01293 | 0.04345 | 23.31 | <2e-16 *** | |
| --- | | | | | |

Does basketball player height affect scoring?

The Collider

$$X \rightarrow Z \leftarrow Y$$

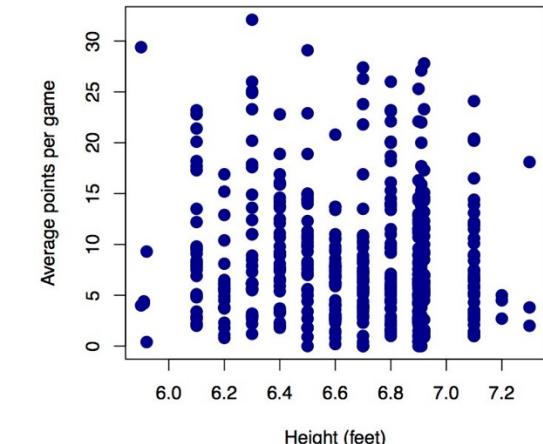
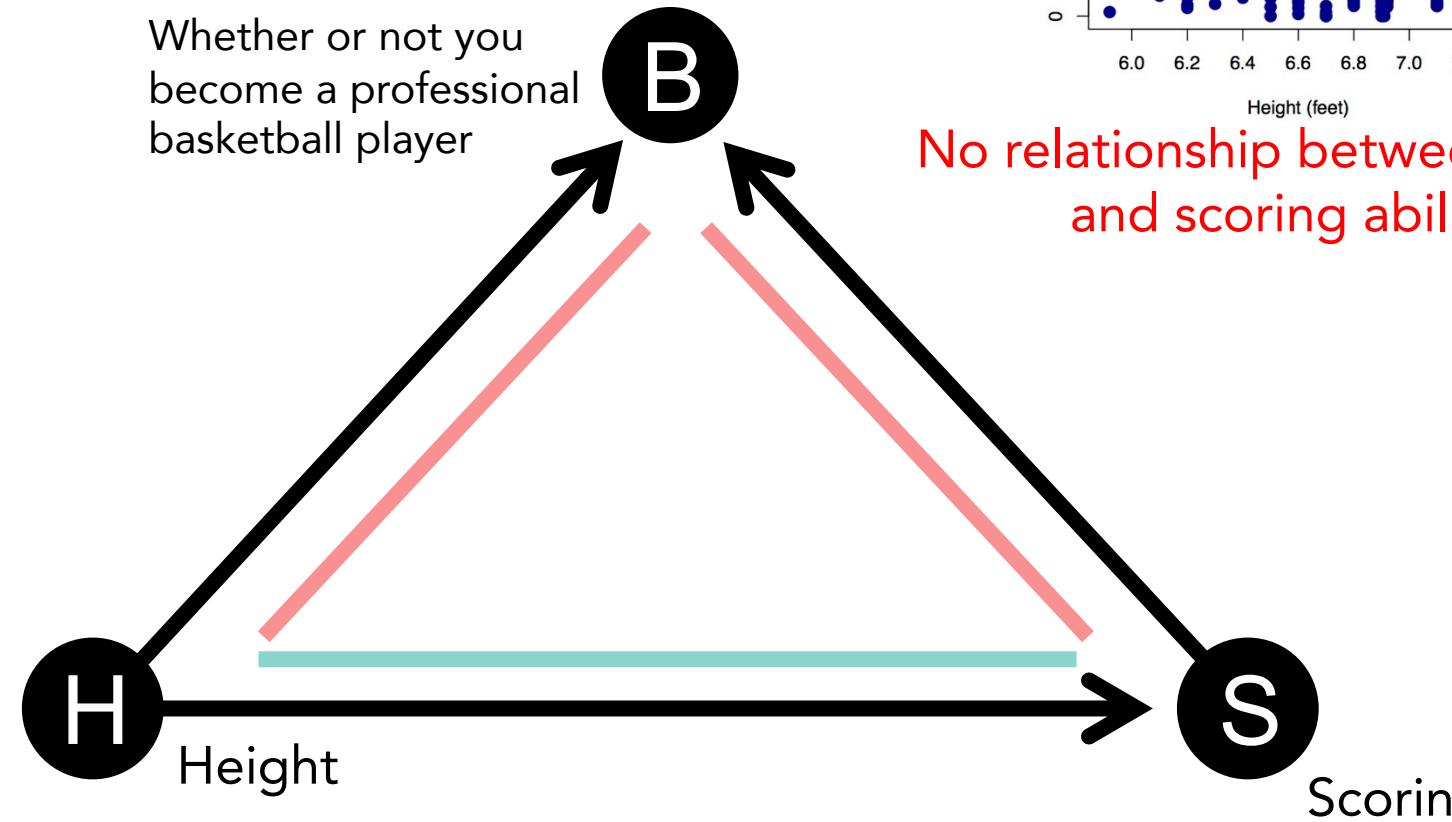
Closed until you
condition on Z



Two paths:

- Path A: $H \rightarrow S$
- Path B: $H \rightarrow B \leftarrow S$

Whether or not you
become a professional
basketball player



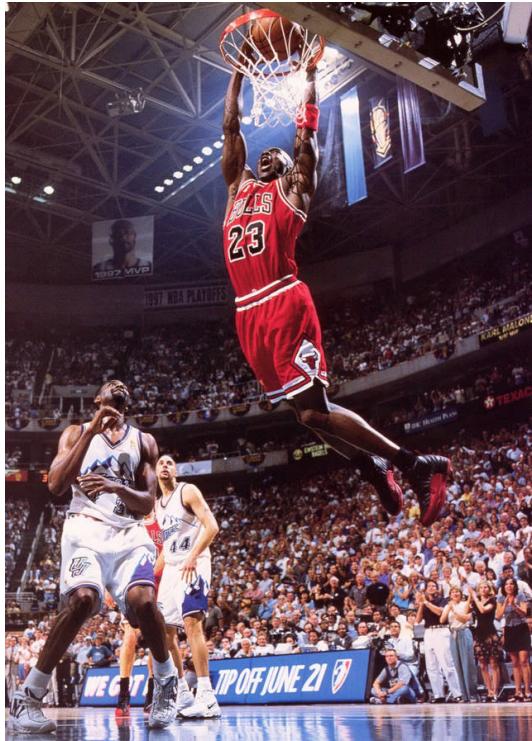
No relationship between height
and scoring ability!

Does basketball player height affect scoring?

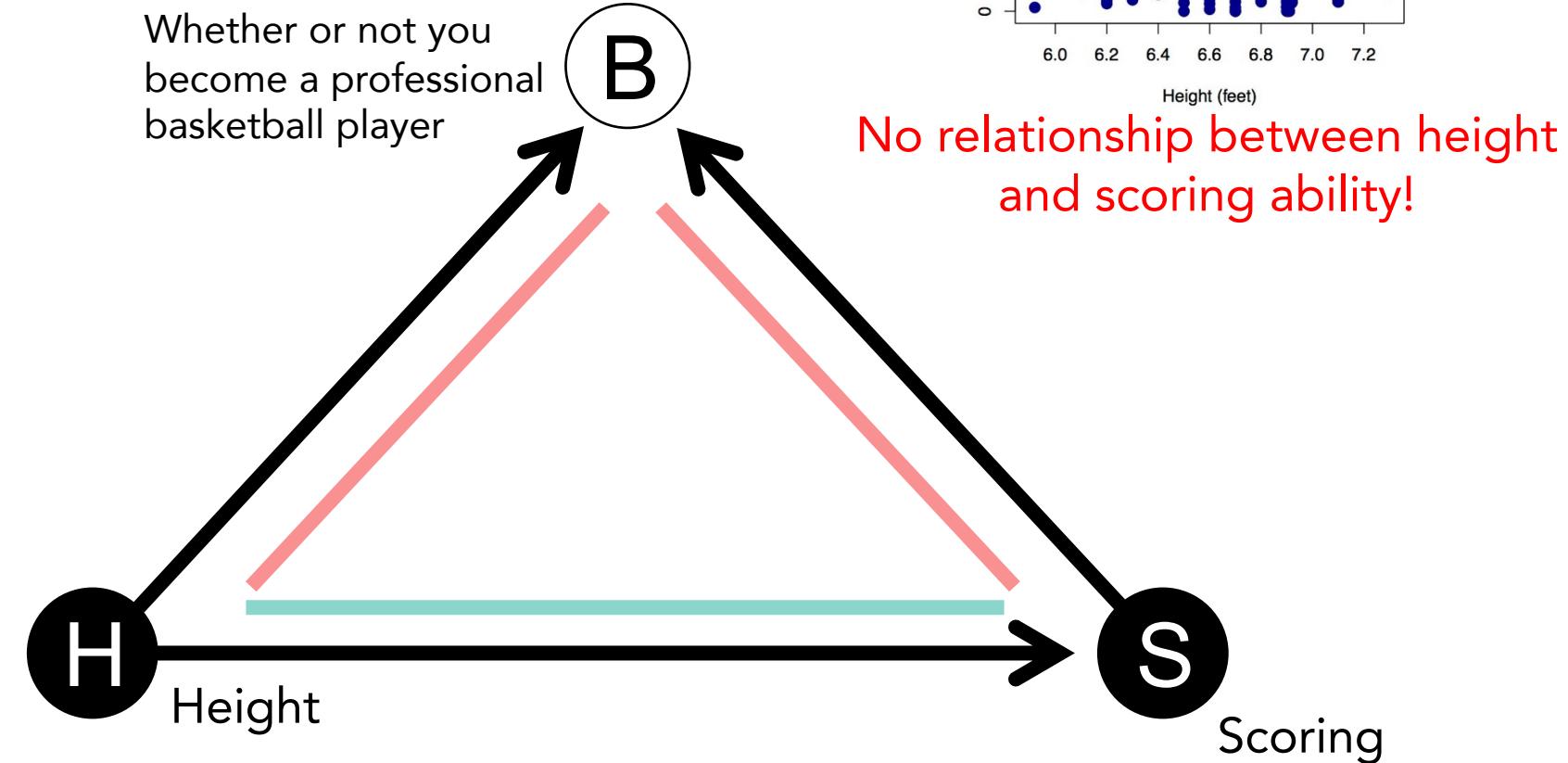
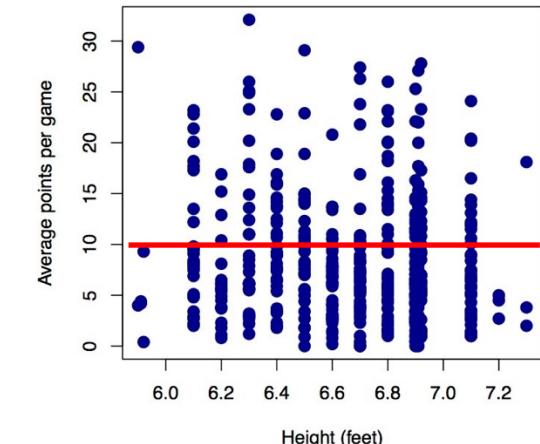
The Collider

$$X \rightarrow Z \leftarrow Y$$

Closed until you
condition on Z



Looking at a subset of B
is like conditioning on B !
 $S \sim H + B$



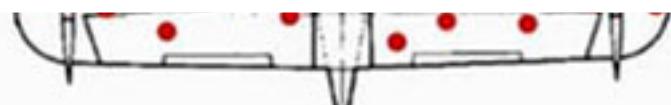
nature > career column > article

CAREER COLUMN | 27 September 2021

Beware survivorship bias in advice on science careers

For objective careers advice, talk to those who left science as well as those who stayed.

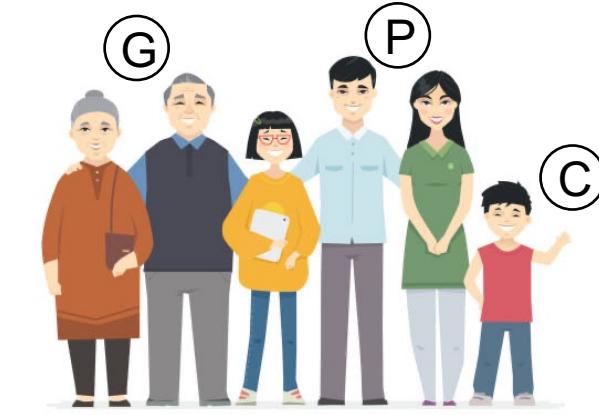
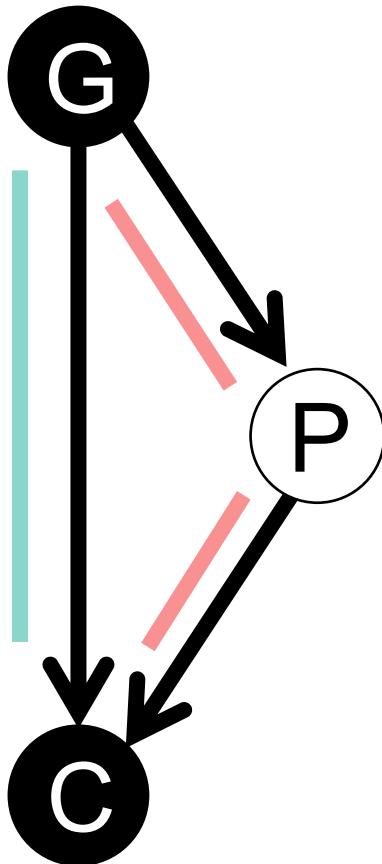
Dave Hemprich-Bennett , Dani Rabaiotti & Emma Kennedy



Chapter 3: Damned if you do.



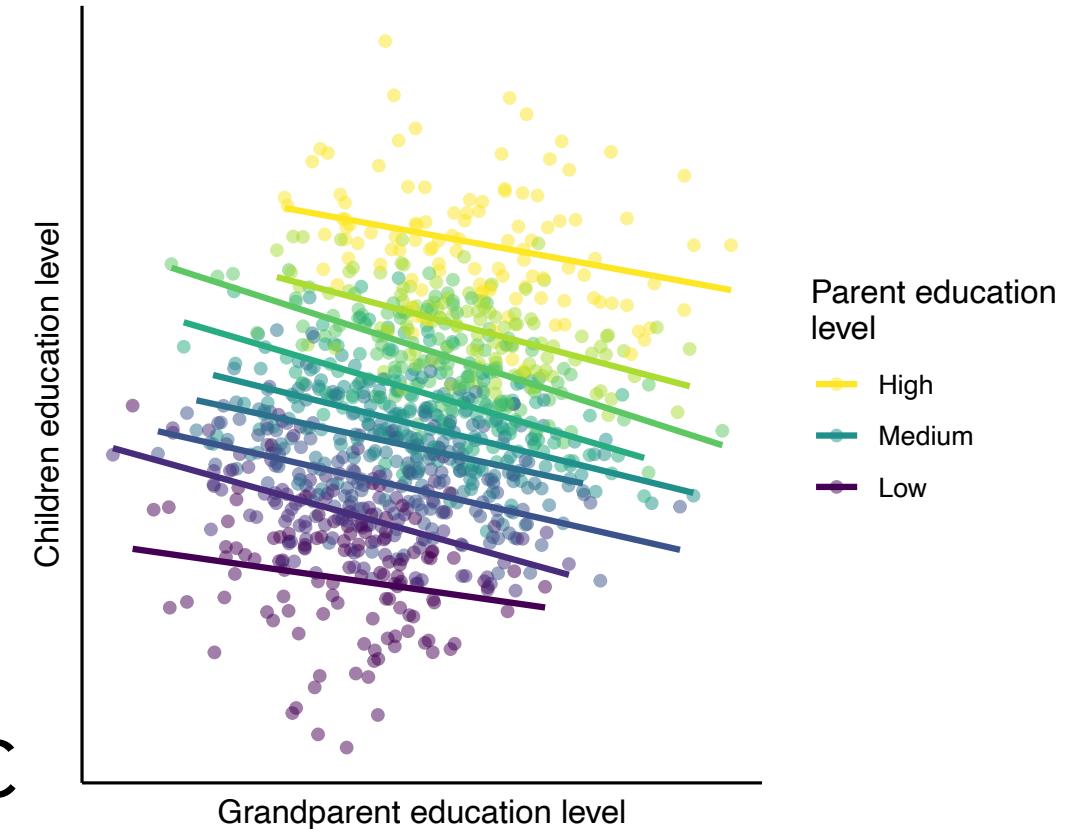
Does your grandparent's education level affect your own?



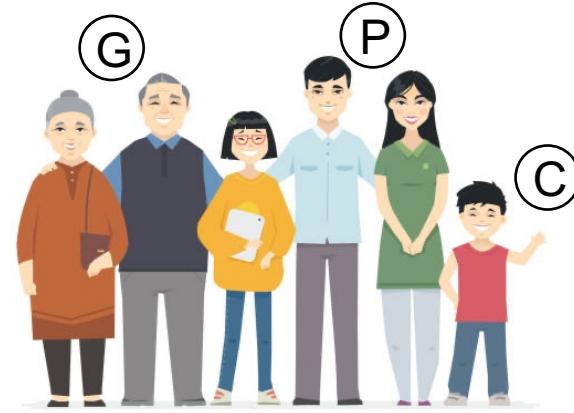
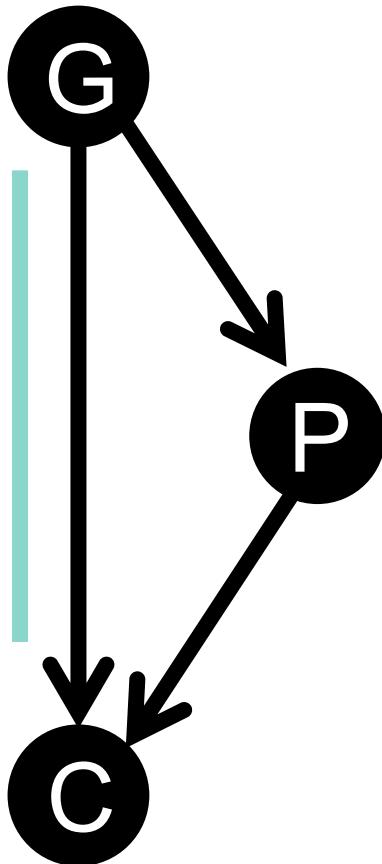
- Two paths:
- Path A: $G \rightarrow C$
 - Path B: $G \rightarrow P \rightarrow C$

The Pipe
 $X \rightarrow Z \rightarrow Y$

Open unless you
condition on Z

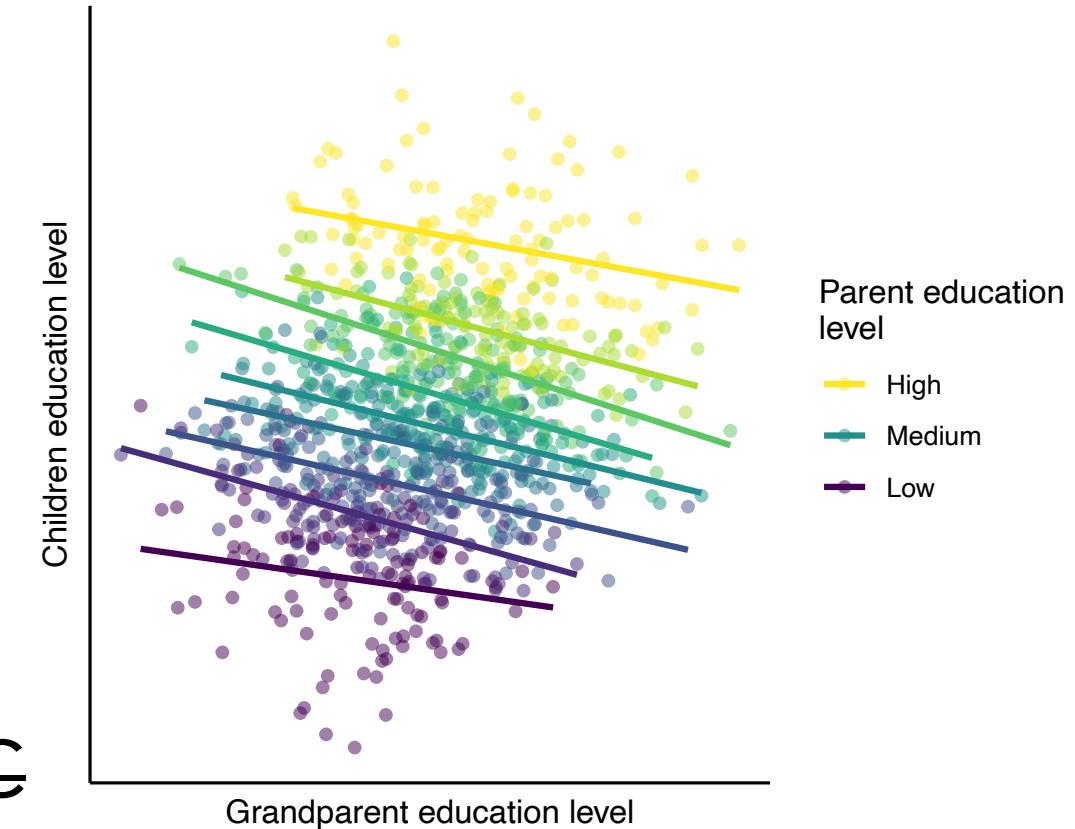


Does your grandparent's education level affect your own?



- Two paths:
- Path A: $G \rightarrow C$
 - Path B: $G \rightarrow P \rightarrow C$

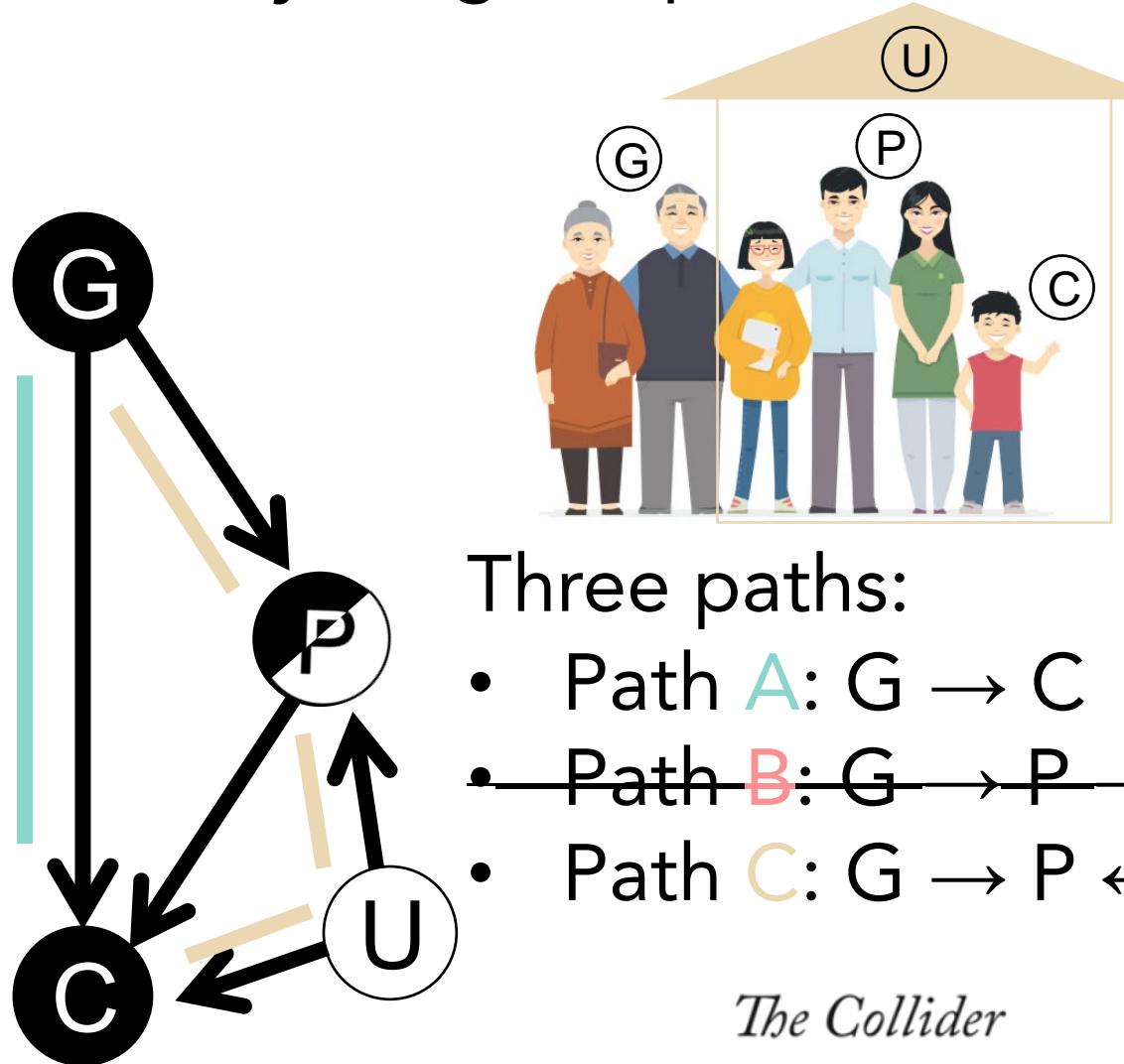
Ideal model?
 $C \sim G + P$



Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|-----------|------------|---------|------------|
| (Intercept) | -0.020116 | 0.055275 | -0.364 | 0.716 |
| G | -1.235624 | 0.053557 | -23.071 | <2e-16 *** |
| P | 2.209555 | 0.022349 | 98.864 | <2e-16 *** |
| G:P | 0.003598 | 0.018804 | 0.191 | 0.848 |

Does your grandparent's education level affect your own?



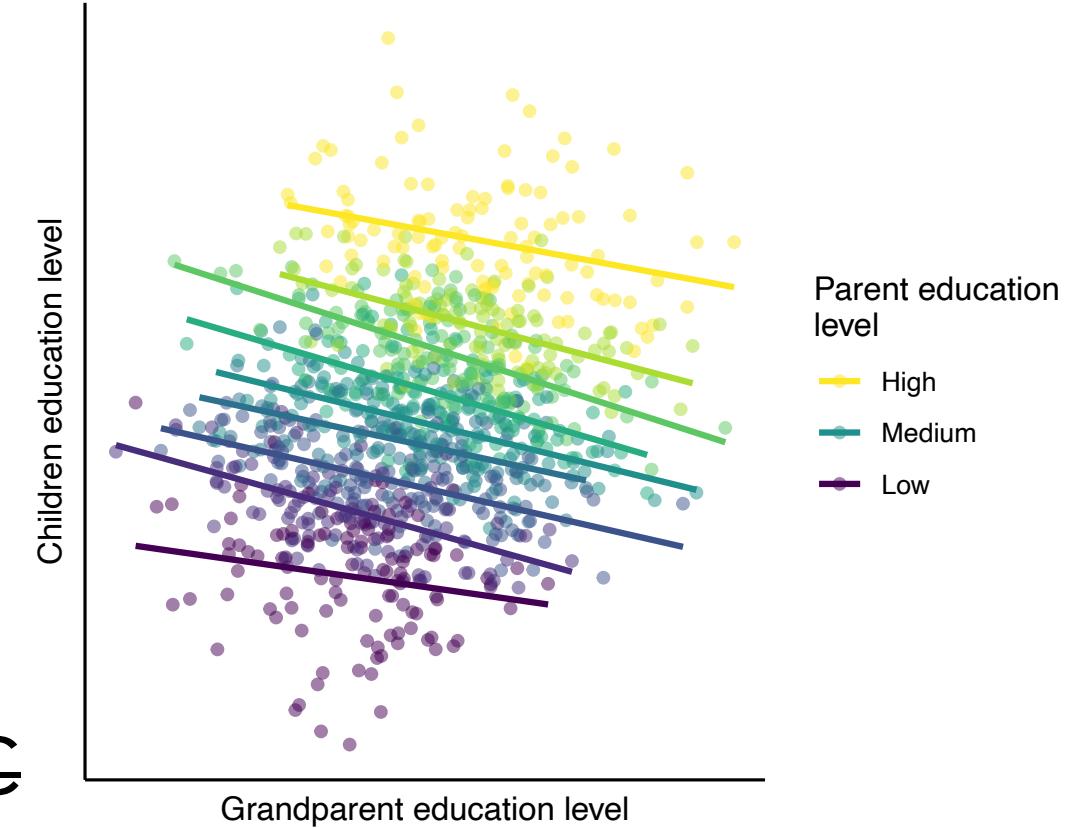
Three paths:

- Path A: $G \rightarrow C$
- Path B: $G \rightarrow P \rightarrow C$
- Path C: $G \rightarrow P \leftarrow U \rightarrow C$

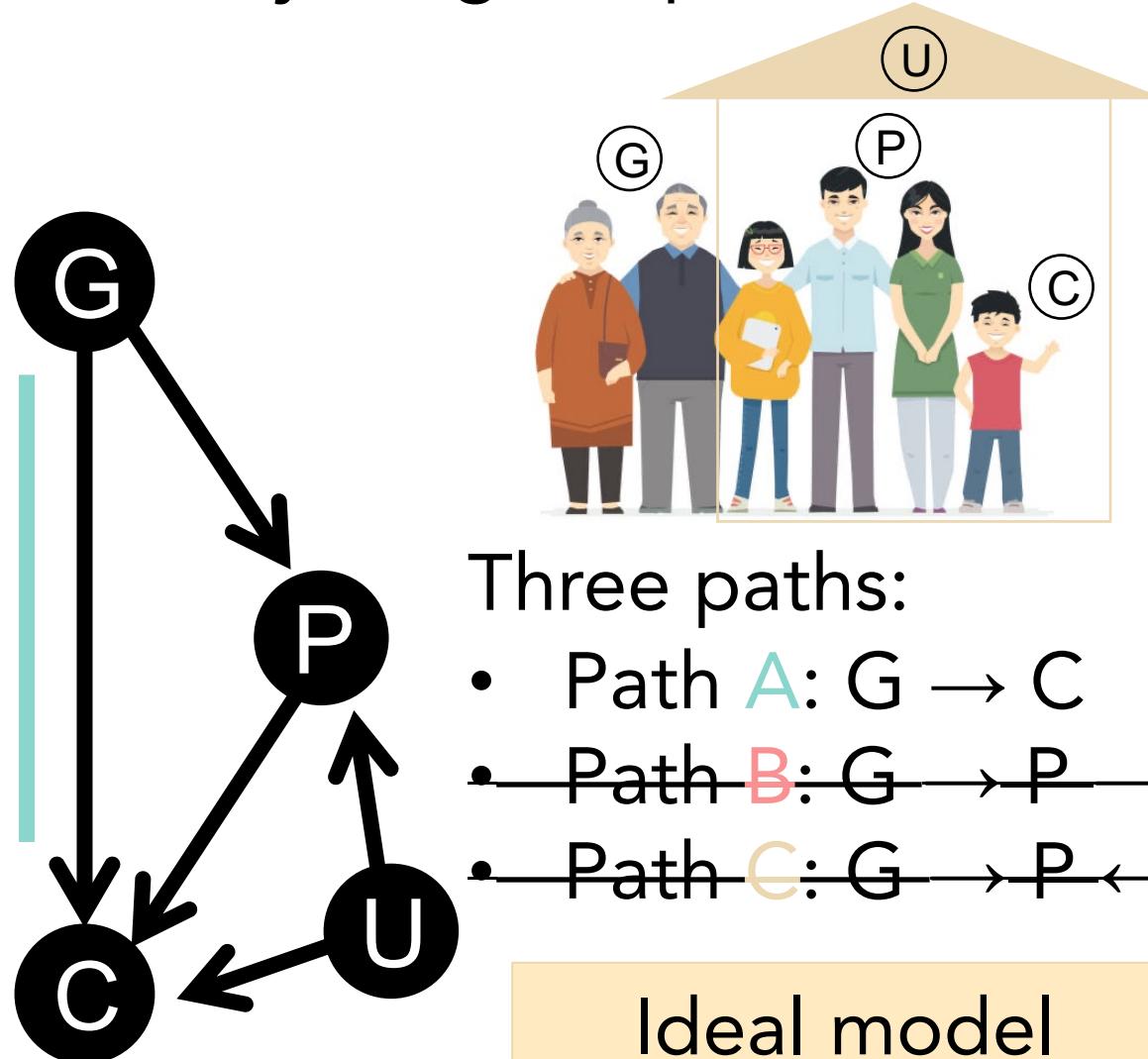
The Collider

$X \rightarrow Z \leftarrow Y$

Closed until you
condition on Z



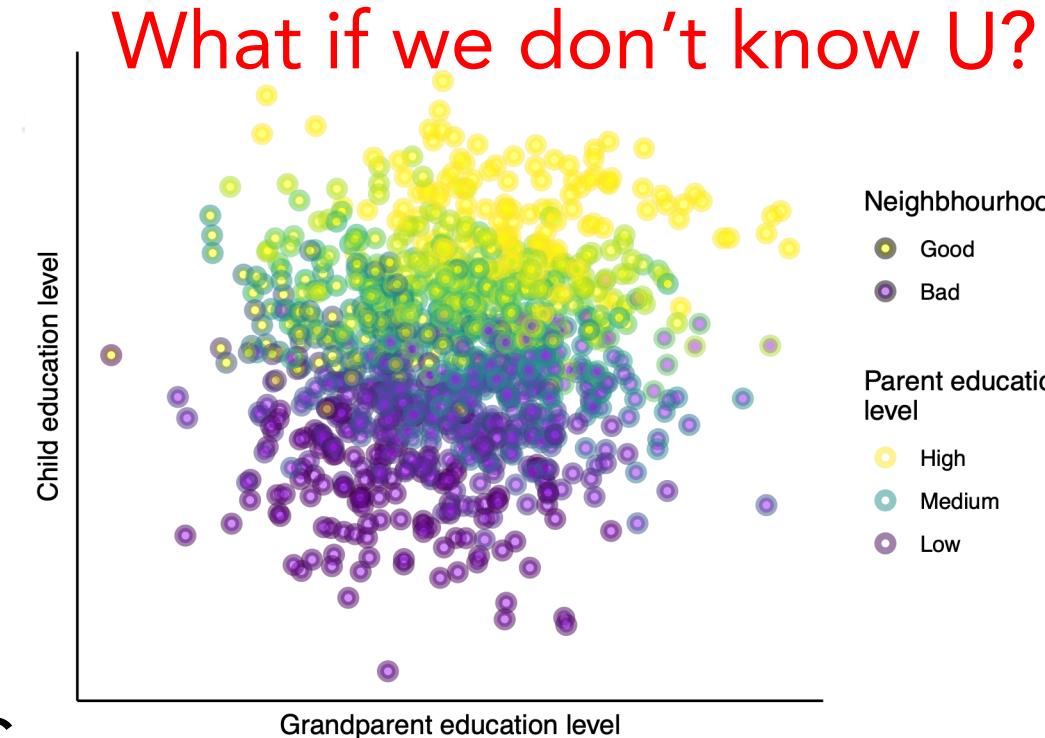
Does your grandparent's education level affect your own?



Three paths:

- Path A: $G \rightarrow C$
- Path B: $G \rightarrow P \rightarrow C$
- Path C: $G \rightarrow P \leftarrow U \rightarrow C$

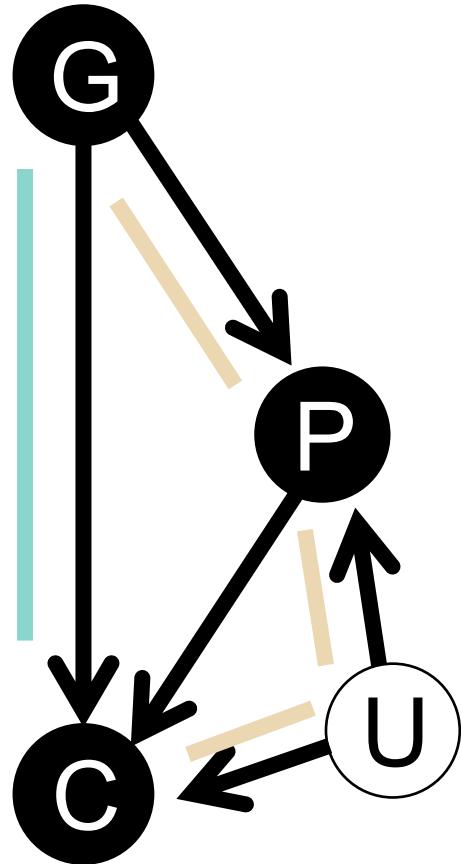
Ideal model
 $C \sim P + C + U$



Coefficients:

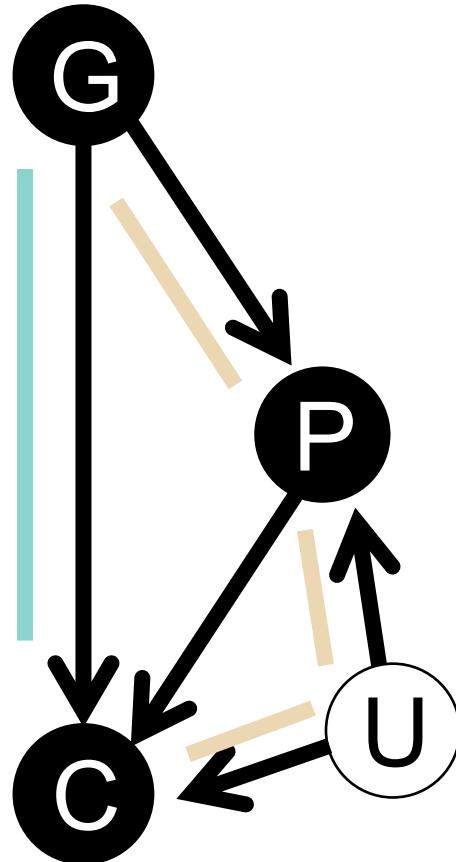
| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|------------|
| (Intercept) | 0.02862 | 0.03217 | 0.890 | 0.374 |
| G | 0.03016 | 0.04438 | 0.679 | 0.497 |
| P | 0.99272 | 0.03178 | 31.241 | <2e-16 *** |
| U | 4.04779 | 0.06930 | 58.412 | <2e-16 *** |

Backdoors are EVERYWHERE in the real world!

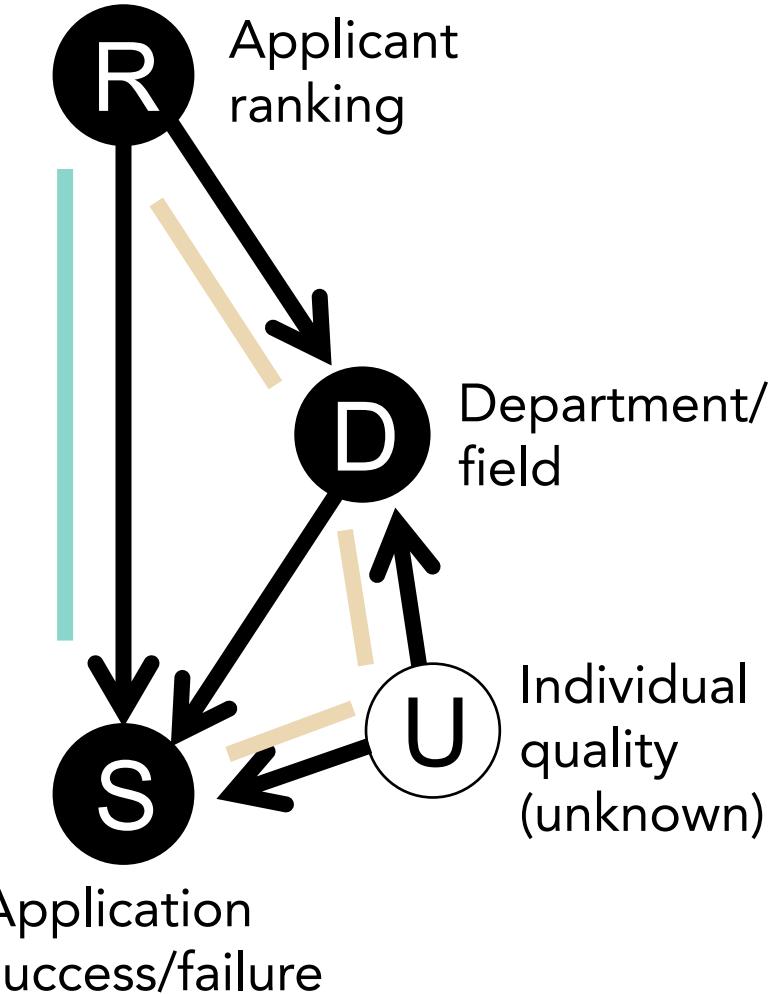


Backdoors are EVERYWHERE in the real world!

E.g., gender bias in applications

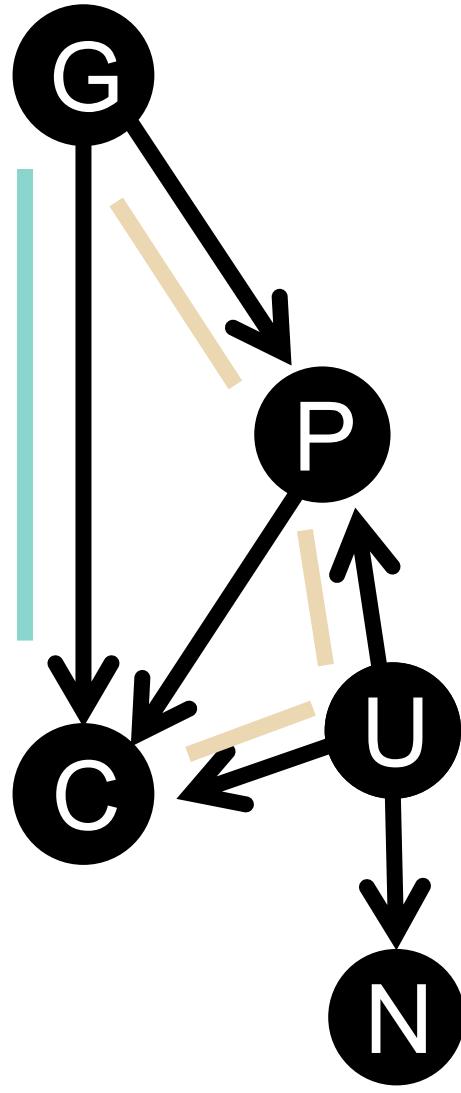


Evidence of biased policing if you don't condition by department, no bias if you do



Backdoors are EVERYWHERE in the real world!

E.g., gender bias in applications



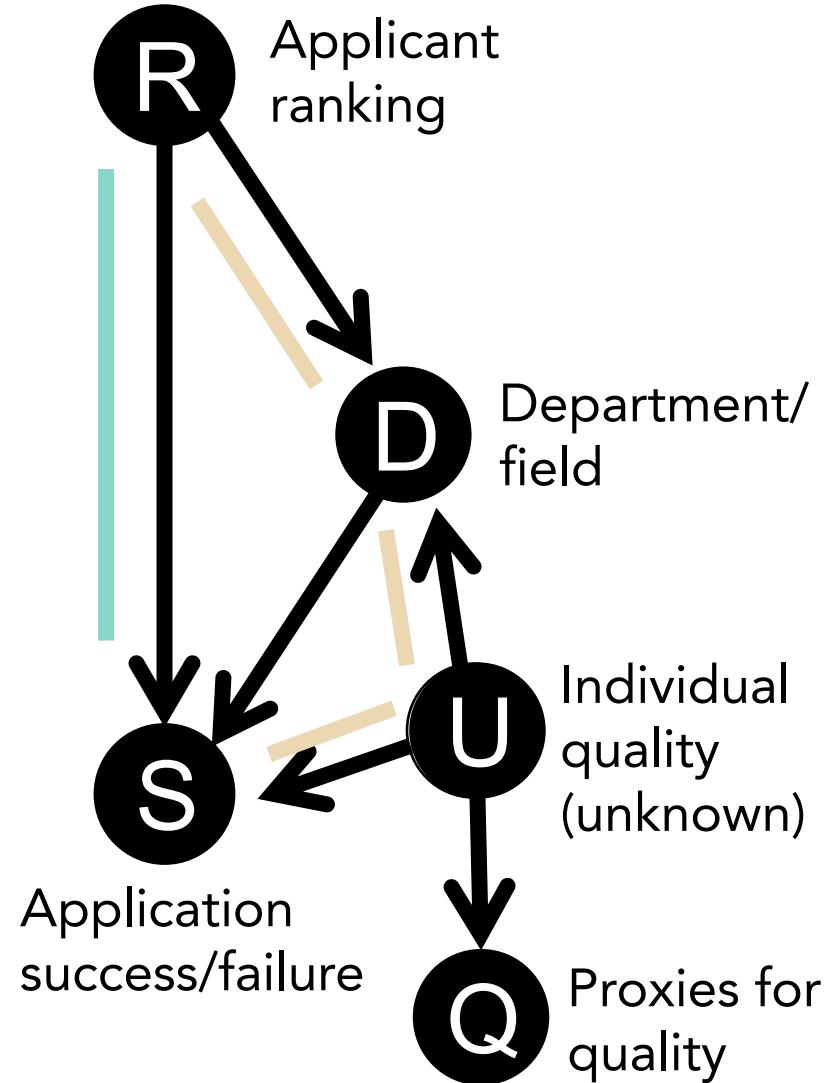
Evidence of biased policing if you don't condition by department, no bias if you do

The Descendant

$$x \rightarrow z \rightarrow y \\ \downarrow \\ a$$

Conditioning on A is like conditioning on Z

Neighbourhood quality,
social capital, etc.



Applicant ranking
Department/field

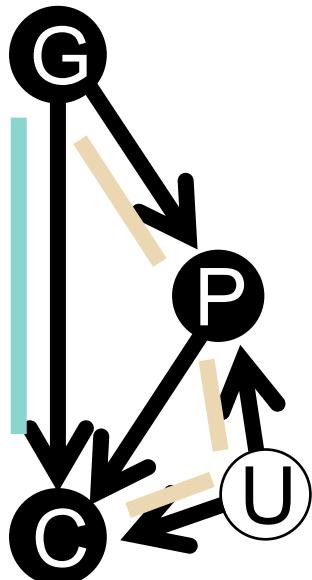
Individual quality (unknown)

Application success/failure

Proxies for quality

Take-aways:

- Use experimental design to exclude confounds' influence(s)
- Use DAGs to form causal assumptions about your system
- When a confound cannot be known, turn to path analysis
 - SEMs: Structural Equation Modelling
 - Bayesian Networks
 - Estimate the influence of unknown/latent parameters
- Never going to be perfect, but always strive for better!



Thanks for listening!

Check out Richard McElreath's YouTube and Statistical Rethinking book if you're interested in learning more!

