

# Kausalität in den Wirtschaftswissenschaften

# Einführung in die Kausalität

Korrelation ist nicht gleich Kausalität

# Einführung in die Kausalität

Korrelation ist nicht gleich Kausalität

Außer, wenn Korrelation gleich Kausalität!

# Einführung in die Kausalität

Korrelation ist nicht gleich Kausalität

Außer, wenn Korrelation gleich Kausalität!

Woher wissen wir, dass X zu Y führt?

# Einführung in die Kausalität

Korrelation ist nicht gleich Kausalität

Außer, wenn Korrelation gleich Kausalität!

Woher wissen wir, dass X zu Y führt?

X führt zu Y, wenn ...

# Einführung in die Kausalität

Korrelation ist nicht gleich Kausalität

Außer, wenn Korrelation gleich Kausalität!

Woher wissen wir, dass X zu Y führt?

X führt zu Y, wenn ...

wir eingreifen und X verändern, alles andere jedoch gleich lassen und ...

# Einführung in die Kausalität

Korrelation ist nicht gleich Kausalität

Außer, wenn Korrelation gleich Kausalität!

Woher wissen wir, dass X zu Y führt?

X führt zu Y, wenn ...

wir eingreifen und X verändern, alles andere jedoch gleich lassen und ...

Y sich danach verändert.

**Achtung:** Y "hört" jedoch nicht unbedingt nur auf X. Es gibt potentiell noch andere Einflüsse, die auf Y einwirken.

# Beispiele für kausale Aussagen

- + Feuerwerkskörper verursachen Lärm, sobald diese angezündet werden
- + Das Krähen des Hahn führt zum Sonnenaufgang
- + Ein Bachelor an der Uni Ulm erhöht ihr späteres Einkommen



# Beispiele für kausale Aussagen

- + Feuerwerkskörper verursachen Lärm, sobald diese angezündet werden
- + Das Krähen des Hahn führt zum Sonnenaufgang
- + Ein Bachelor an der Uni Ulm erhöht ihr späteres Einkommen

**Kausalität = Korrelation + zeitliche Abfolge + kein Scheinzusammenhang**

Woher wissen Sie, dass Sie die richtigen Faktoren betrachten (d.h. keinen Scheinzusammenhang)?

# Beispiele für kausale Aussagen

- + Feuerwerkskörper verursachen Lärm, sobald diese angezündet werden
- + Das Krähen des Hahn führt zum Sonnenaufgang
- + Ein Bachelor an der Uni Ulm erhöht ihr späteres Einkommen

**Kausalität = Korrelation + zeitliche Abfolge + kein Scheinzusammenhang**

Woher wissen Sie, dass Sie die richtigen Faktoren betrachten (d.h. keinen Scheinzusammenhang)?

Hier benötigen Sie ein Modell!

# Datengenerierungsprozess

Um zu verstehen, wie wir aus experimentellen Daten oder aus Beobachtungsdaten *kausale* Zusammenhänge ableiten können müssen wir uns folgenden Fragen widmen:

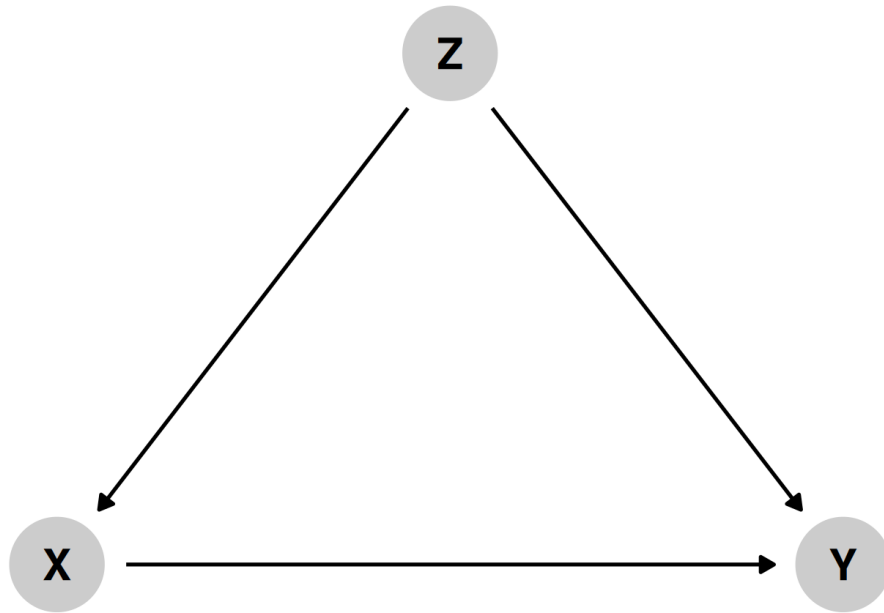
- + Woher stammen unsere Daten (Experiment oder Beobachtungsdaten)?
- + Welcher **Datengenerierungsprozess** hat diese Daten erzeugt?
- + Können wir auf diesen Datengenerierungsprozess einwirken? (insbesondere durch ein Experiment)
- + Gibt es zufällige Elemente in unseren Daten die wir analysieren könnten?

# Kausale Graphen (DAGs)

Um uns den Datengenerierungsprozess vor Augen zu führen sollten wir diesen modellieren:

- Hierzu verwenden wir die **directed acyclic graphs (DAGs)**, welche vor allem von Judea Pearl entwickelt wurden
- In diesen DAGs fließt Kausalität (acyclic) nur in eine Richtung (falls Rückwärtskausalität oder Simultanität vorhanden ist sind andere Modelle besser geeignet)
- DAGs präsentieren Kausalität als Alternativszenario ("counterfactuals")
  - Ein kausaler Effekt wird als Vergleich zweier Alternativen definiert:
    - Der Zustand der stattgefunden hat mit der Intervention
    - Der Zustand der nicht stattgefunden hat ohne die Intervention (das "counterfactual")

# Wie sehen diese DAGs aus?



- + **Directed:** Jede Node hat einen Pfeil der zu einer anderen Node zeigt
- + **Acyclic:** Pfeile haben nur eine Richtung, sie können nicht zu einer Node zurück

Daher können Sie DAGs folgendermaßen interpretieren:

- + Grafische Repräsentation ihres Modells
- + Jeder Pfeil zeigt eine kausale Verbindung einer Variablen zur nächsten an (Richtung der Kausalität)
- + Dort wo es keine Pfeile zwischen den Variablen gibt vermuten Sie keinen kausalen Zusammenhang (Annahmen ersichtlich)

# Wie sehen diese DAGs aus?

Doch woher kommt das DAG?

# Wie sehen diese DAGs aus?

Doch woher kommt das DAG?

- + Eine Sehr gute Frage!
- + Es entsteht aus ihren Erkenntnissen aus der Literatur, den eigenen Hypothesen, ökonomische Theorie, ihren eigenen Beobachtungen ...

# Zusammenhängende Ansätze zur Kausalität

Zwei unterschiedliche, aber verwandte, Ansätze über Kausalität nachzudenken:



# Zusammenhängende Ansätze zur Kausalität

Zwei unterschiedliche, aber verwandte, Ansätze über Kausalität nachzudenken:

## Directed Acyclic Graphs (DAGs)

- + Grafische Modelle
- + Kausalität fließt immer in eine Richtung und wird durch Pfeile verdeutlicht
- + Keine Rückwärtskausalität oder Simultanität abbildbar
- + Betrachtet Alternativszenarien
- + Do-Calculus im Hintergrund (Fancy Mathe)

Verwenden wir in dieser Veranstaltung!

# Zusammenhängende Ansätze zur Kausalität

Zwei unterschiedliche, aber verwandte, Ansätze über Kausalität nachzudenken:

## Directed Acyclic Graphs (DAGs)

- + Grafische Modelle
- + Kausalität fließt immer in eine Richtung und wird durch Pfeile verdeutlicht
- + Keine Rückwärtskausalität oder Simultanität abbildbar
- + Betrachtet Alternativszenarien
- + Do-Calculus im Hintergrund (Fancy Mathe)

## Potential Outcomes Modell (auch Rubin Kausalmodell genannt)

- + Betrachtet Alternativszenarien
- + Betrachten von durchschnittlichen Effekten über mehrere Individuen oder Gruppen (ATE)
- + Unterschied zwischen Gruppen wenn eine Intervention getätigt vs. nicht getätigt wurde
- + Annahmen: Stable Unit Treatment Value Assumption (SUTVA)

Verwenden wir in dieser Veranstaltung!

# Zusammenhängende Ansätze zur Kausalität

Zwei unterschiedliche, aber verwandte, Ansätze über Kausalität nachzudenken:

## Directed Acyclic Graphs (DAGs)

- + Grafische Modelle
- + Kausalität fließt immer in eine Richtung und wird durch Pfeile verdeutlicht
- + Keine Rückwärtskausalität oder Simultanität abbildbar
- + Betrachtet Alternativszenarien
- + Do-Calculus im Hintergrund (Fancy Mathe)

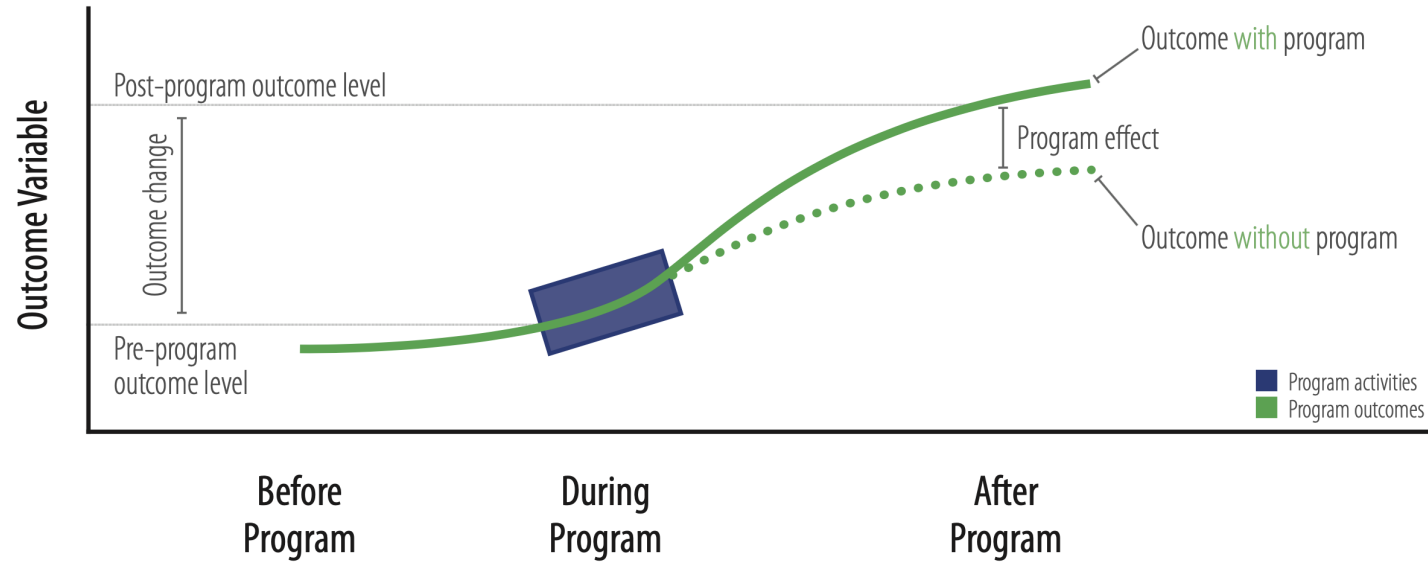
## Potential Outcomes Modell (auch Rubin Kausalmodell genannt)

- + Betrachtet Alternativszenarien
- + Betrachten von durchschnittlichen Effekten über mehrere Individuen oder Gruppen (ATE)
- + Unterschied zwischen Gruppen wenn eine Intervention getätigt vs. nicht getätigt wurde
- + Annahmen: Stable Unit Treatment Value Assumption (SUTVA)

**Verwenden wir in dieser Veranstaltung!**

Insbesondere die DAGs helfen uns den zugrunde liegenden Datengenerierungsprozess zu modellieren.

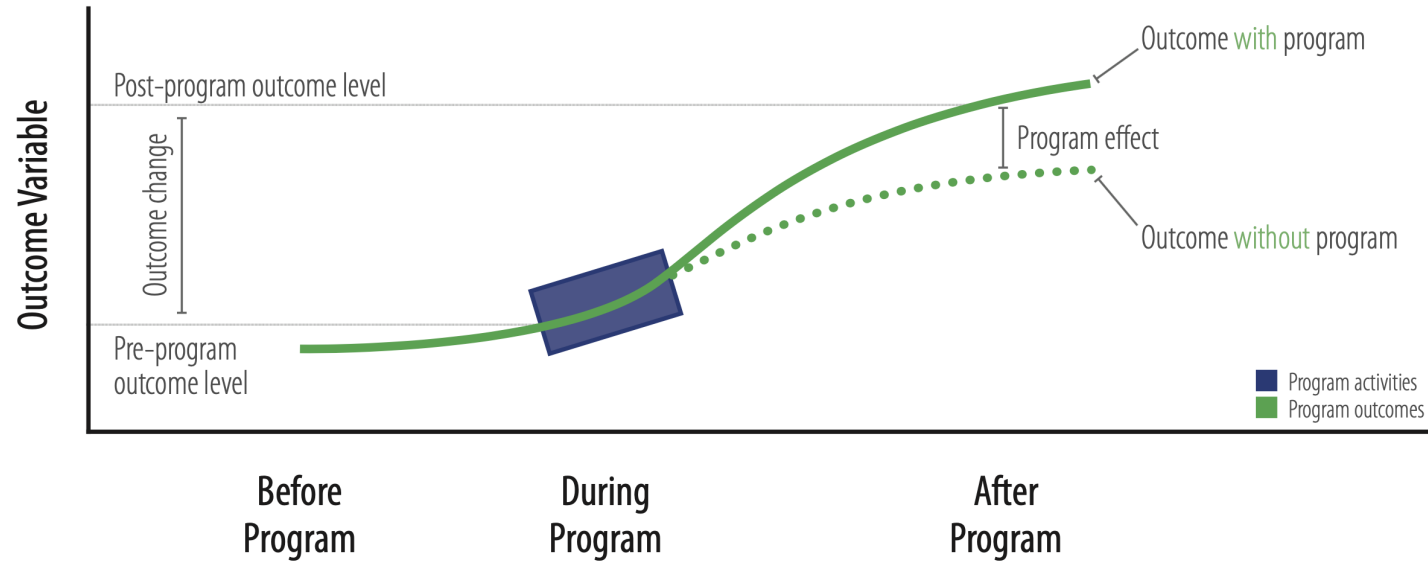
# Wie groß war der Effekt einer Intervention?



Quelle: Andrew Heiss, Programm evaluation (<https://evalf20.classes.andrewheiss.com/content/01-content/>)

Wie messen wir den Effekt einer Intervention?

# Wie groß war der Effekt einer Intervention?



Quelle: Andrew Heiss, Programm evaluation (<https://evalf20.classes.andrewheiss.com/content/01-content/>)

Wie messen wir den Effekt einer Intervention?

Mit einem Experiment

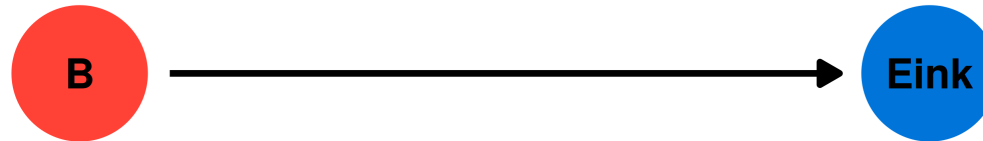
Mit Beobachtungsdaten

# Die Auswirkung von Bildung auf das Einkommen

Erhöht ein Jahr mehr Bildung (z.B. durch einen Bachelor an der Uni Ulm) ihr späteres Einkommen?!

# Die Auswirkung von Bildung auf das Einkommen

Erhöht ein Jahr mehr Bildung (z.B. durch einen Bachelor an der Uni Ulm) ihr späteres Einkommen?!



# Beispielhaftes DAG für unsere Kausalitätsfrage

Mögliche Faktoren, die die Bildungsentscheidung von Personen beeinflussen und deren Einkommen:

- + **Schritt Eins:** Mögliche Variablen sammeln
- + **Schritt Zwei:** Vereinfachen
- + **Schritt Drei:** DAG zeichnen
- + **Schritt Vier:** Modell erstellen welches getestet werden kann



# Schritt eins: Mögliche Variablen sammeln

Bildung (treatment) → Einkommen (outcome)

Welche Variablen spielen möglicherweise noch eine Rolle?

# Schritt eins: Mögliche Variablen sammeln

Bildung (treatment) → Einkommen (outcome)

Welche Variablen spielen möglicherweise noch eine Rolle?

- + (Unbeobachtbare) Fähigkeiten
- + Haushaltseinkommen
- + Sozialer Status
- + Bildung der Eltern
- + Schulpflicht
- + Netzwerk

# Schritt zwei: Vereinfachen

Bildung (treatment) → Einkommen (outcome)

Wie kann ich diese Variablen vereinfachen?

- + (Unbeobachtbare) Fähigkeiten
- + Haushaltseinkommen
- + Sozialer Status
- + Bildung der Eltern
- + Schulpflicht
- + Netzwerk

# Schritt zwei: Vereinfachen

Bildung (treatment) → Einkommen (outcome)

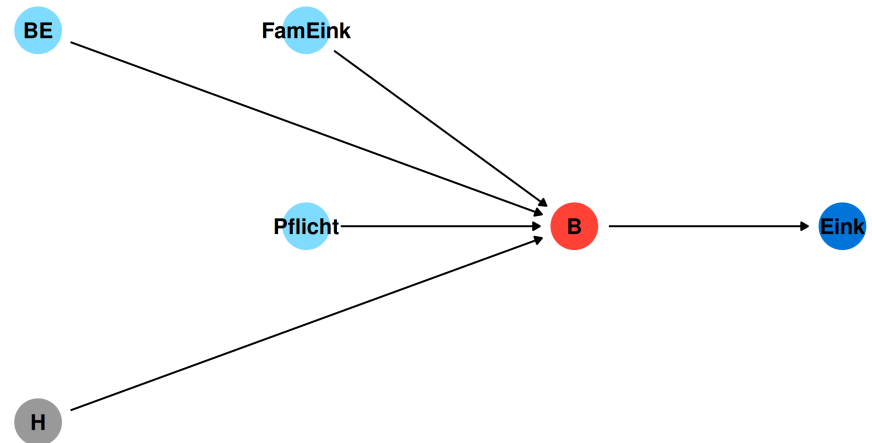
Wie kann ich diese Variablen vereinfachen?

- + (Unbeobachtbare) Fähigkeiten → Hintergrund
- + Haushaltseinkommen
- + Sozialer Status → Hintergrund
- + Bildung der Eltern
- + Schulpflicht
- + Netzwerk

# Schritt drei: DAG zeichnen

Hintergrund, Haushaltseinkommen, die Bildung der Eltern und Schulpflicht

→ Determinanten ihrer **Bildung!**



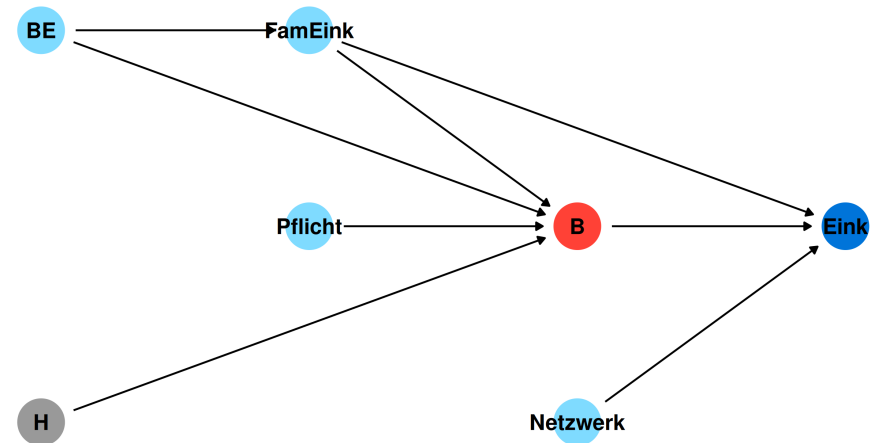
# Schritt drei: DAG zeichnen

Haushaltseinkommen und ihr Netzwerk

→ Determinanten für ihr **Einkommen!**

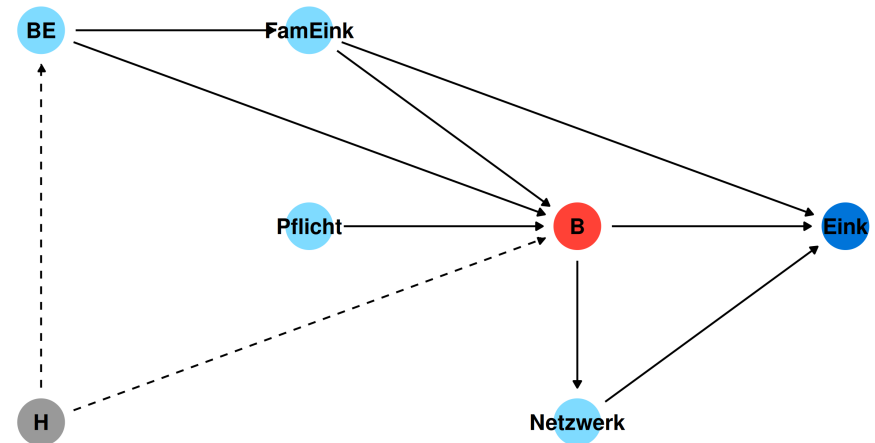
**Weiterhin:**

- ✚ Bildung der Eltern ist entscheidend für das Haushaltseinkommen.



# Schritt drei: DAG zeichnen

- + Ihr Netzwerk wird auch von ihrer Bildungsentscheidung beeinflusst
- + Hintergrund hat sicherlich auch die Bildungsentscheidung ihrer Eltern beeinflusst!
- + Weiterhin ist dieser **Hintergrund nicht beobachtbar** (daher grau) und sollte mit gestrichelten Linien verdeutlicht werden.

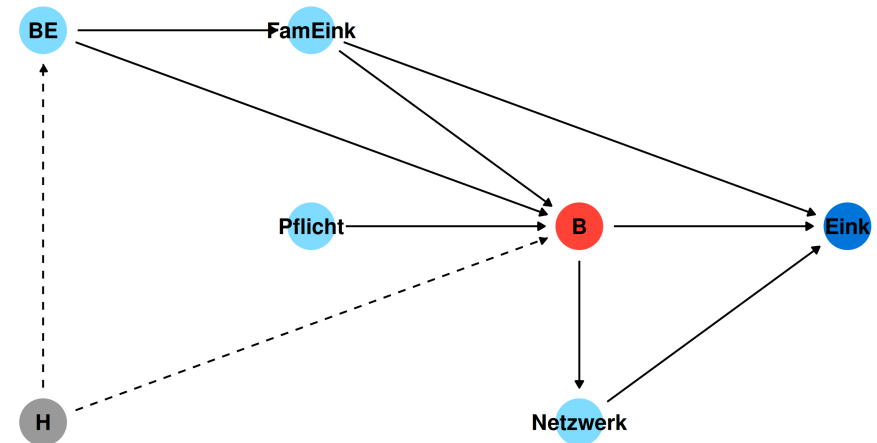


# Schritt vier: Identifikation

Nun haben Sie die einzelnen Variablen und deren Zusammenhang aufgezeigt! Alle diese Nodes sind miteinander korreliert

Wie können wir nun den Effekt von **B** auf **Eink** extrahieren?

Um den Effekt von **B** auf **Eink** zu identifizieren müssen Sie diesen von den anderen umgebenden Effekten isolieren!





# Schritt vier: Identifikation

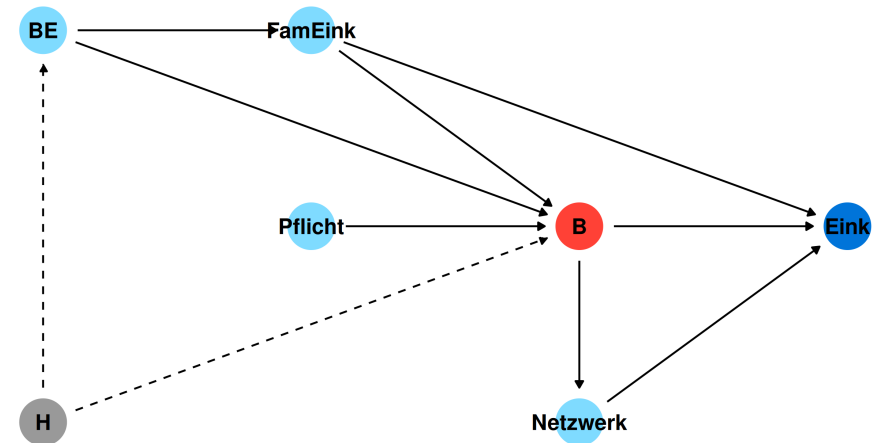
Nun haben Sie die einzelnen Variablen und deren Zusammenhang aufgezeigt! Alle diese Nodes sind miteinander korreliert

Wie können wir nun den Effekt von **B** auf **Eink** extrahieren?

Um den Effekt von **B** auf **Eink** zu identifizieren müssen Sie diesen von den anderen umgebenden Effekten isolieren!

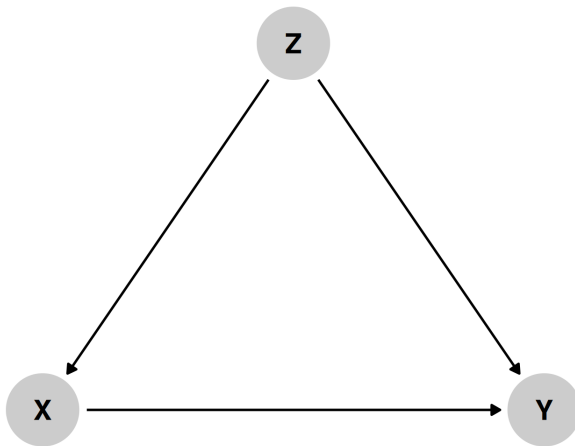
Eine Möglichkeit wäre es auf die umgebenden Effekte in einer Regression zu "kontrollieren".

Sollten wir nun einfach auf alle Variablen unseres DAGs kontrollieren?



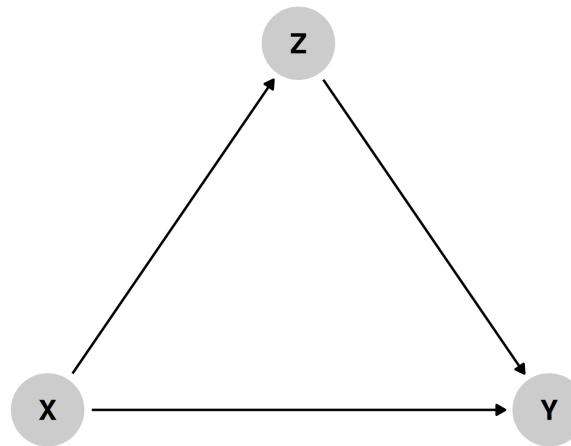
# Welche Arten von Zusammenhang gibt es im DAG?

Confounder



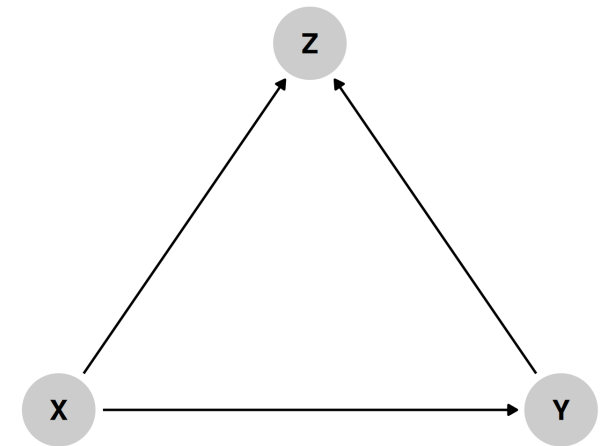
Hier handelt es sich um einen gemeinsamen Ursprung

Mediator



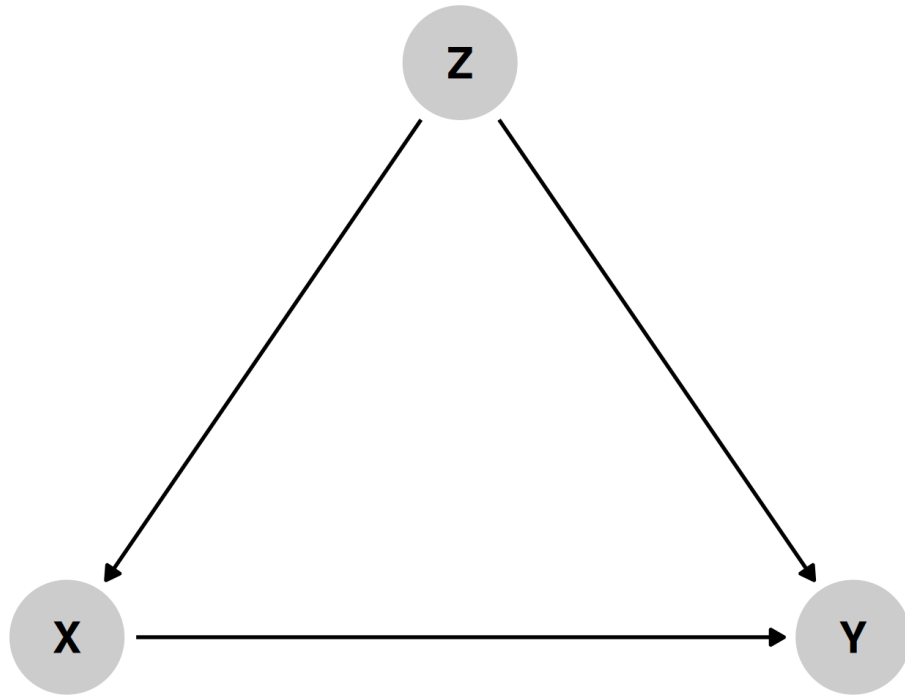
Hier handelt es sich um einen Mittler des Gesamteffekts

Collider



Hier handelt es sich um Selektion / Endogenität

# Confounder



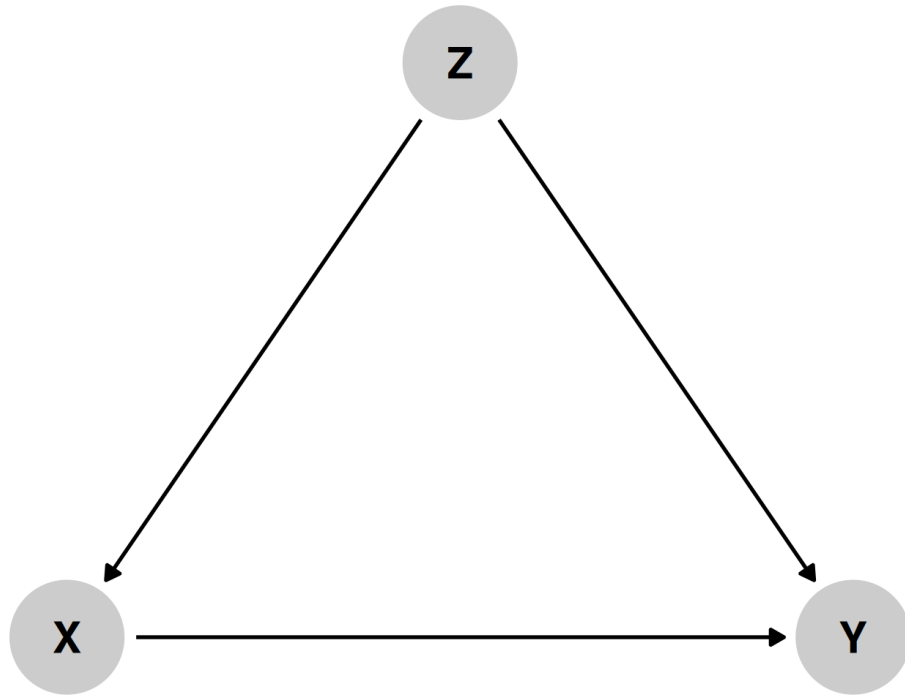
X führt zu Y

+ Dies sehen wir am *kausalen* Pfad  $X \rightarrow Y$

Z führt dazu, dass sich sowohl X als auch Y verändern

Wir wollen den *kausalen* Effekt von X auf Y isolieren.

# Confounder



X führt zu Y

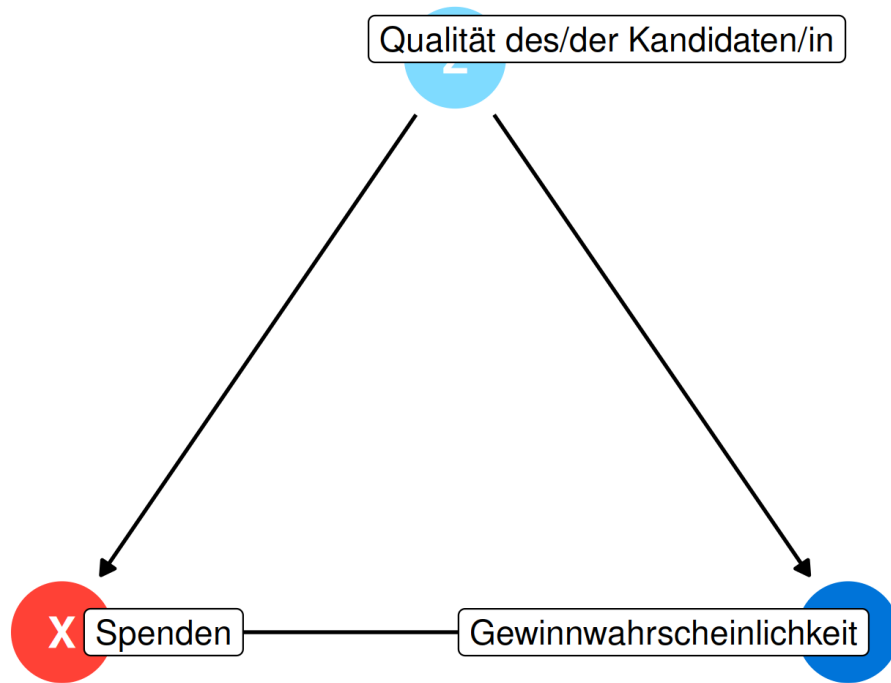
+ Dies sehen wir am *kausalen* Pfad  $X \rightarrow Y$

Z führt dazu, dass sich sowohl X als auch Y verändern

Wir wollen den *kausalen* Effekt von X auf Y isolieren.

- + Wir sprechen hier davon das Z ein *confounder* des kausalen Zusammenhangs zwischen  $X \rightarrow Y$  ist
- + Der Confounder ist eine dritte Variable, welche sowohl X, als auch Y beeinflusst
- + Oft lesen Sie auch von der *backdoor* Variablen Z
- + Der *backdoor* Pfad  $X \leftarrow Z \rightarrow Y$  generiert eine Scheinkorrelation zwischen X und Y
- + Eine *backdoor* offen zu lassen generiert Bias da die Beziehung zwischen X und Y nicht isoliert wurde!

# Confounder



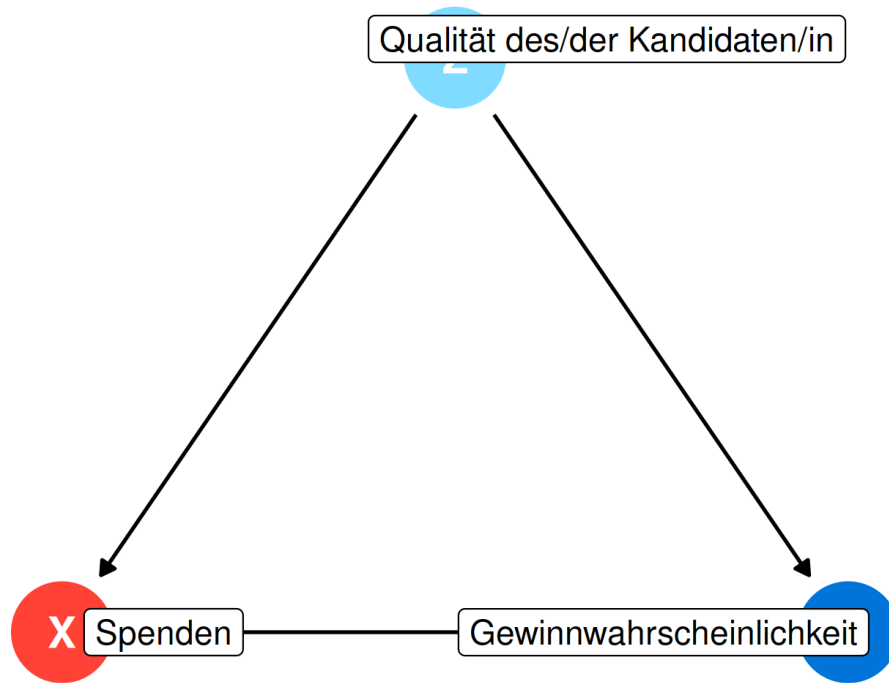
Pfade zwischen **Spenden** und **Gewinnwahrscheinlichkeit**

Spenden  $\rightarrow$  Gewinnwahrscheinlichkeit

Spenden  $\leftarrow$  Qualität des/der Kandidaten/in  $\rightarrow$   
Gewinnwahrscheinlichkeit

**Qualität des/der Kandidaten/in ist eine *backdoor***

# Confounder



Pfade zwischen **Spenden** und **Gewinnwahrscheinlichkeit**

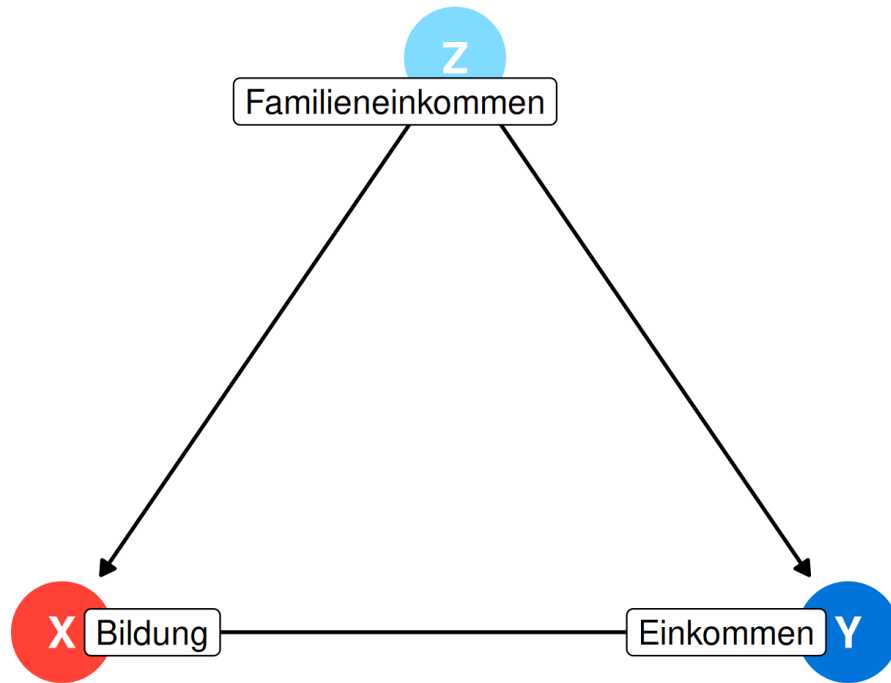
$\text{Spenden} \rightarrow \text{Gewinnwahrscheinlichkeit}$

$\text{Spenden} \leftarrow \text{Qualität des/der Kandidaten/in} \rightarrow \text{Gewinnwahrscheinlichkeit}$

**Qualität des/der Kandidaten/in** ist eine *backdoor*

Sie schließen die *backdoor* indem sie auf **Z** kontrollieren

# Confounder



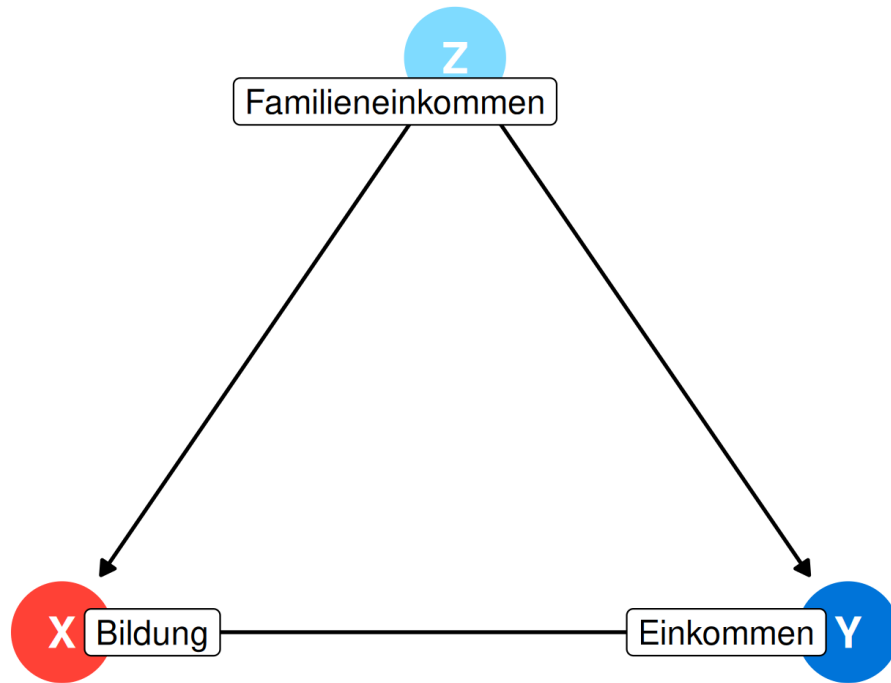
Pfade zwischen **Bildung** und **Einkommen**

Bildung  $\rightarrow$  Einkommen

Bildung  $\leftarrow$  Familieneinkommen  $\rightarrow$  Einkommen

**Familieneinkommen ist eine *backdoor***

# Confounder



Pfade zwischen **Bildung** und **Einkommen**

Bildung  $\rightarrow$  Einkommen

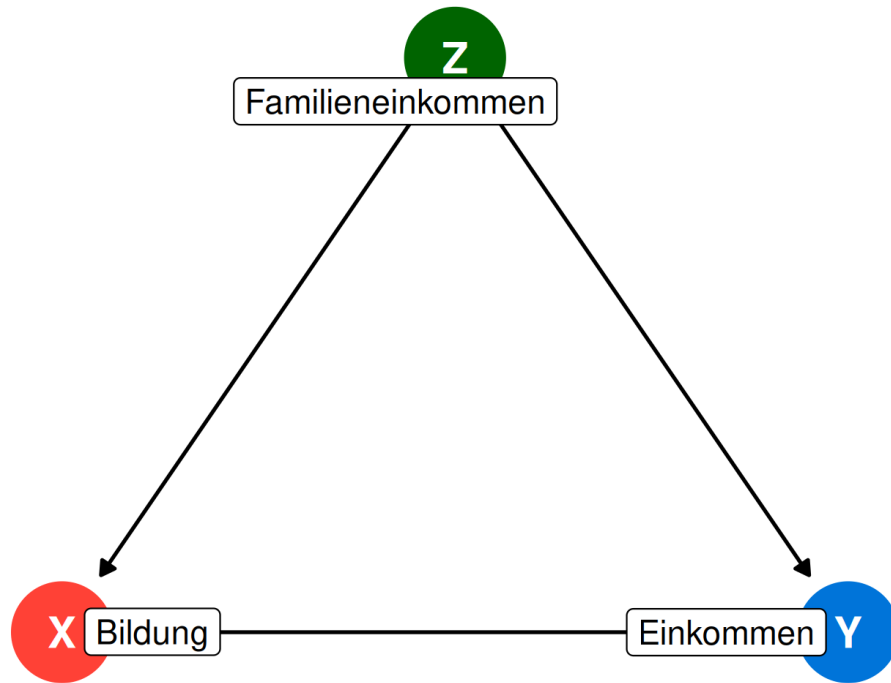
Bildung  $\leftarrow$  Familieneinkommen  $\rightarrow$  Einkommen

**Familieneinkommen ist eine *backdoor***

Sie schließen die *backdoor* indem sie auf **Z** kontrollieren



# Türen schließen



Dadurch das Sie auf das Familieneinkommen kontrollieren:

- + Eliminieren Sie den Effekt des Familieneinkommens auf die Bildung
- + Eliminieren Sie den Effekt des Familieneinkommens auf das spätere Einkommen

Zusammenhang des verbleibenden Effekts der Bildung auf den verbleibenden Effekt des Einkommens ist unser **kausaler Effekt** von Bildung auf Einkommen.

→ Durch die Kontrolle vergleichen wir hier Personen, wie wenn diese das gleiche Familieneinkommen hätten

# Kontrollieren innerhalb einer Regression

Eine Möglichkeit auf Variablen zu "kontrollieren" ist mittels einer multiplen linearen Regression:

$$\text{Einkommen} = \beta_0 + \beta_1 * \text{Bildung} + \beta_2 * \text{Familieneinkommen} + \epsilon$$

Etwas weitreichendere Methoden (diese wollen wir jedoch nicht näher besprechen):

- + Matching
- + Stratifizierung
- + Synthetische Kontrollgruppen

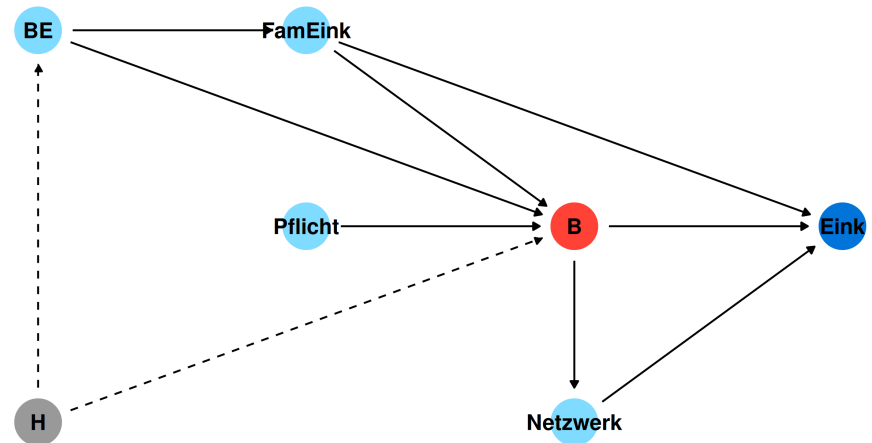
# Backdoor Pfade schließen

Bildung  $\rightarrow$  Einkommen

Bildung  $\leftarrow$  Familieneinkommen  $\rightarrow$  Einkommen

Bildung  $\leftarrow$  elterliche Bildung  $\rightarrow$  Familieneinkommen  $\rightarrow$   
Einkommen

Bildung  $\leftarrow$  Hintergrund  $\rightarrow$  elterliche Bildung  $\rightarrow$   
FamEink  $\rightarrow$  Einkommen



# Kontrollieren innerhalb einer Regression

Das Familieneinkommen liegt auf jeder dieser *backdoor* Pfade

Kontrollieren wir auf das Familieneinkommen, so können wir die jeweiligen *backdoor* Pfade schließen

Durch die Kontrolle auf das Familieneinkommen erfüllen wir das *Backdoor Kriterium*.

$$Einkommen = \beta_0 + \beta_1 * Bildung + \beta_2 * Familieneinkommen + \epsilon$$

In diesem Fall wäre der Koeffizient von ( $\beta_1$ ) kausal zu interpretieren!

# *Backdoor Pfade* schließen

Doch was ist, wenn unser DAG nicht komplett war?

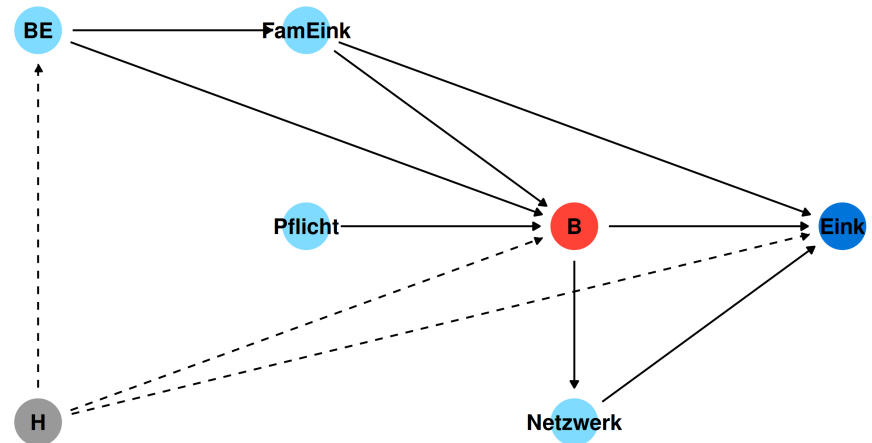
# Backdoor Pfade schließen

Doch was ist, wenn unser DAG nicht komplett war?

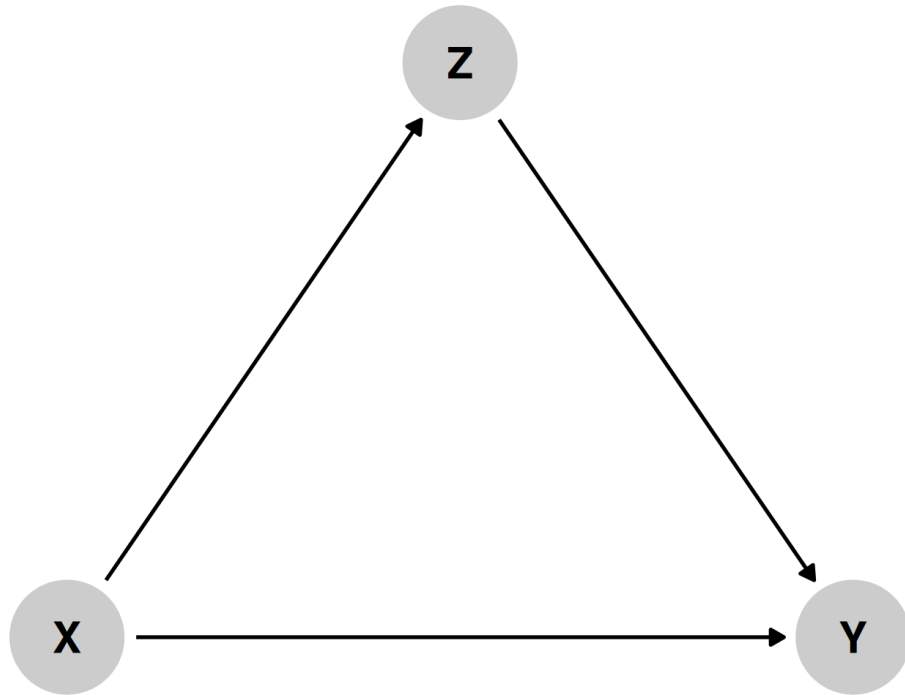
Z.B. könnte der **Hintergrund** einer Person auch einen direkten Effekt auf dessen Einkommen haben:

Bildung  $\leftarrow$  Hintergrund  $\rightarrow$  Einkommen

Hier können wir die *Backdoor* nicht schließen und dürften den Koeffizienten ( $\beta_1$ ) **nicht** kausal interpretieren!



# Mediator

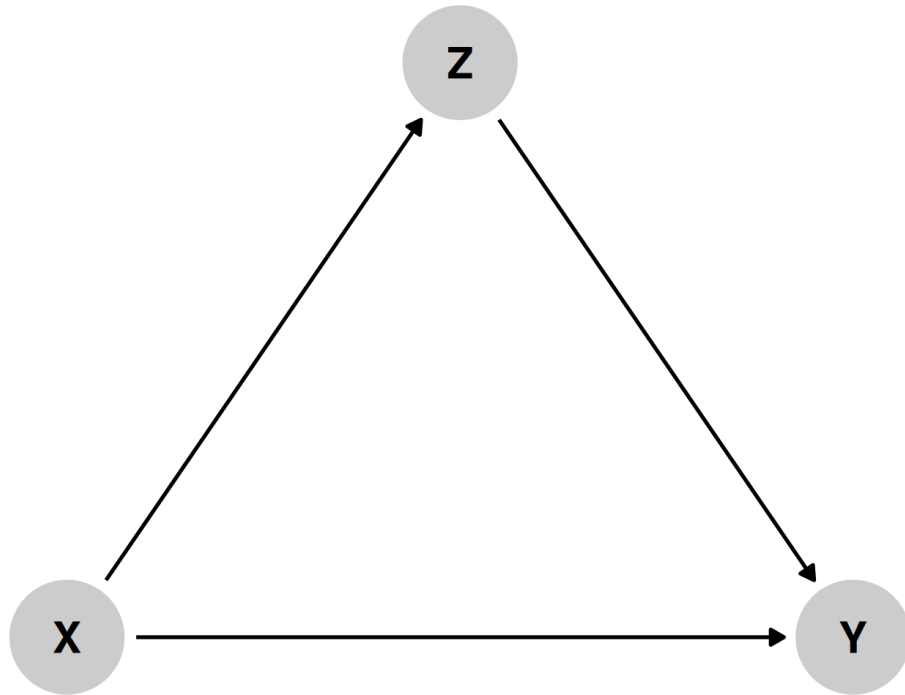


X führt zu Y

X führt zu Z, welches wiederum zu Y führt

Sollten wir für Z kontrollieren?

# Mediator



X führt zu Y

X führt zu Z, welches wiederum zu Y führt

Sollten wir für Z kontrollieren?

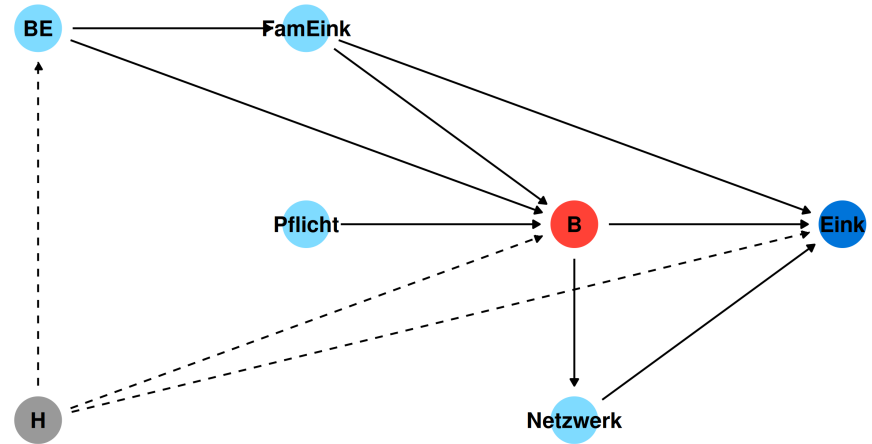
**Nein** → Dies würde zu einer Überanpassung des Modells führen!

Ein Teil des Effekts von X auf Y würde damit außer Acht gelassen.

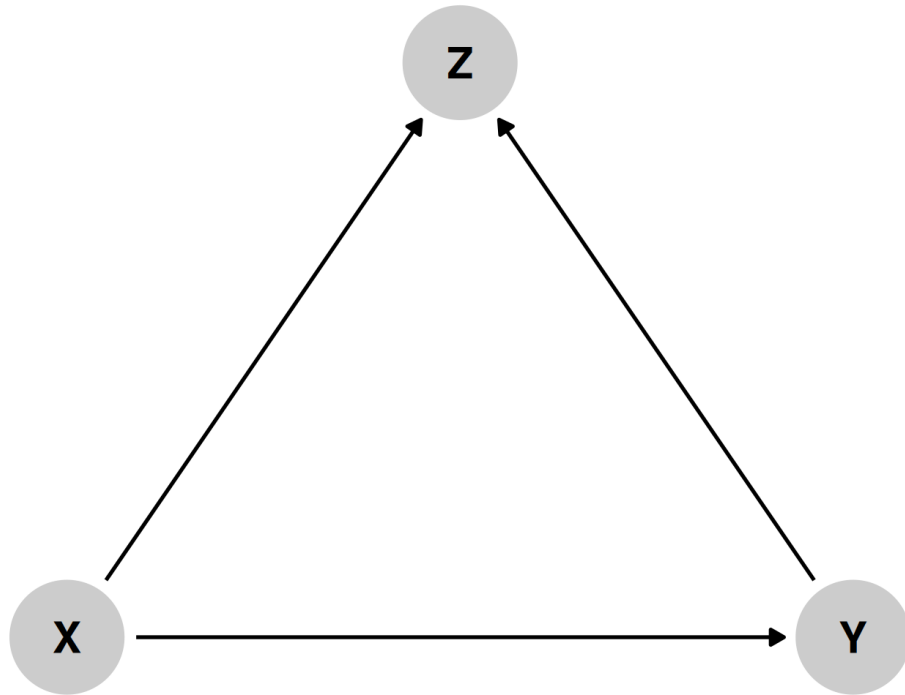


# Überanpassung

Sollten wir in unserem Modell für Netzwerk kontrollieren?



# Collider

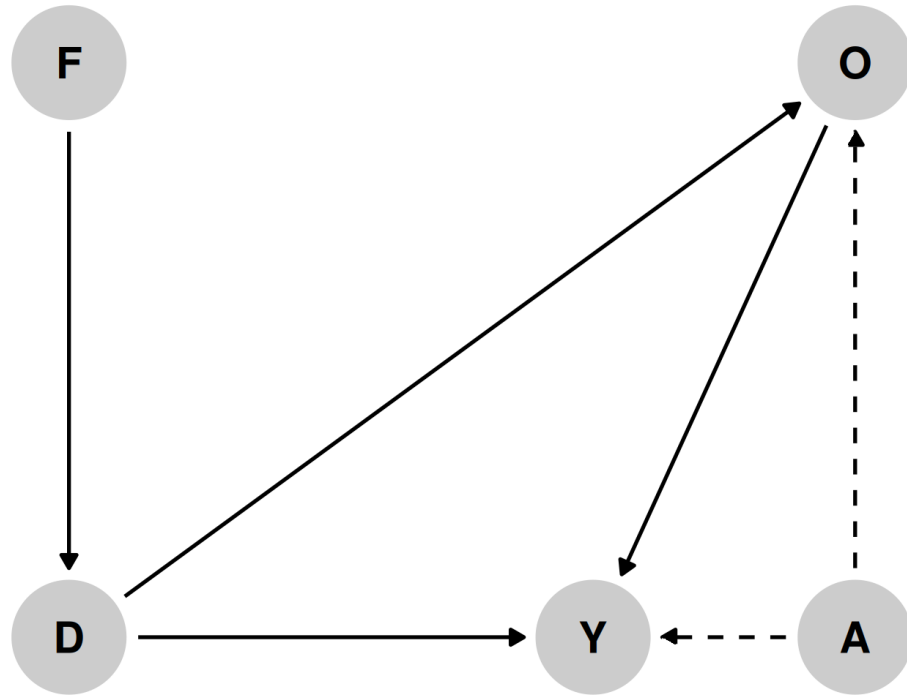


X führt zu Y

X führt zu Z und Y führt zu Z

Sollten wir für Z kontrollieren?

# Beispiel zu Collider



Kritiker bemängeln, dass Google Frauen systematisch schlechter bezahlt.

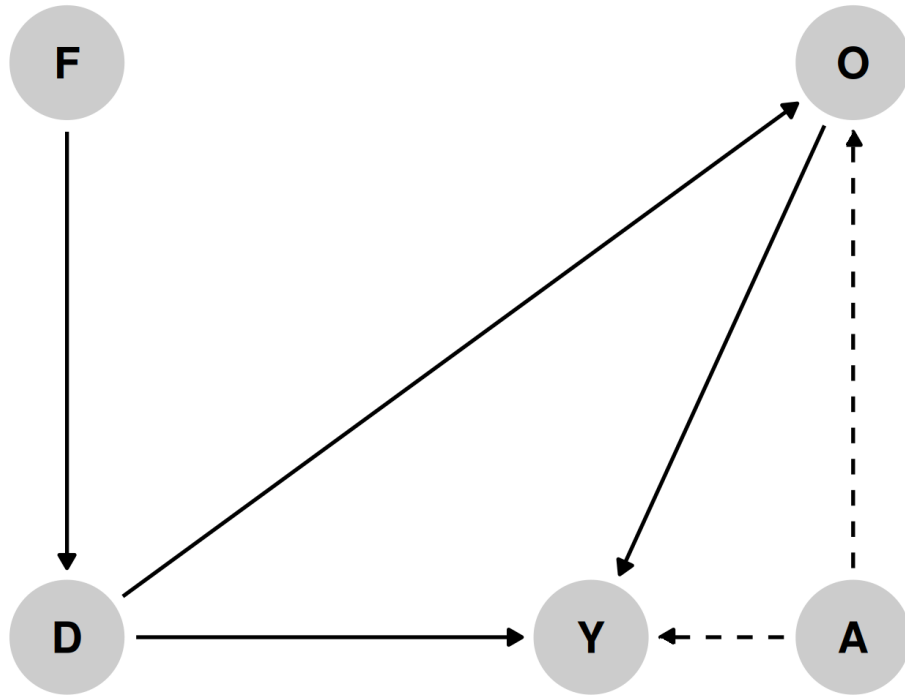
Google entgegnet: Wenn wir den Ort, die Dauer, den Jobtitel, das Managementlevel und die Performance mit in die Betrachtung einfließen lassen, dann verdienen Frauen gleich viel wie Männer.

Doch was wenn Diskriminierung durch Auswahl bestimmter beruflicher Bereiche stattfindet?

Das DAG dazu sehen Sie links!

- + F = Frau
- + D = Diskriminierung
- + O = Beruf
- + Y = Einkommen
- + A = Fähigkeiten

# Beispiel zu Collider



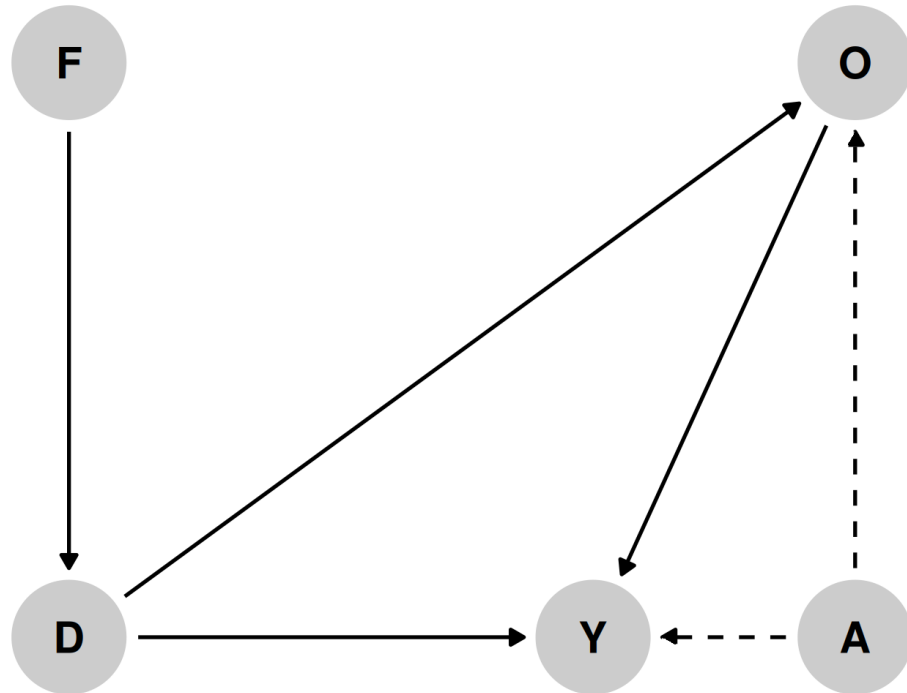
- + F = Frau
- + D = Diskriminierung
- + O = Beruf
- + Y = Einkommen
- + A = Fähigkeiten

$D \rightarrow O \rightarrow Y$

$D \rightarrow O \leftarrow A \rightarrow Y$

Warum ist der Beruf (O) ein Collider?

# Beispiel zu Collider



- + F = Frau
- + D = Diskriminierung
- + O = Beruf
- + Y = Einkommen
- + A = Fähigkeiten

$D \rightarrow O \rightarrow Y$

$D \rightarrow O \leftarrow A \rightarrow Y$

Warum ist der Beruf (O) ein Collider?

Wenn wir auf den Beruf (O) kontrollieren, öffnen wir den zweiten Kanal, da O als Collider fungiert.

Collider blocken immer die *backdoor*, wenn auf den Collider kontrolliert wird, dann öffnen wir die *backdoor*!

# Beispiel zu Collider

Wir wollen uns die Auswirkungen des Collider Bias näher in einer Simulation anschauen:

```
#Aus dem Buch Causal Inference: The Mixtape
#https://colab.research.google.com/github/scunning1975/mixtape_learnr/blob/main/R_ipynb/Directed_
tb <- tibble(
  female = ifelse(runif(10000)>=0.5,1,0),
  ability = rnorm(10000),
  discrimination = female,
  occupation = 1 + 2*ability + 0*female - 2*discrimination + rnorm(10000),
  wage = 1 - 1*discrimination + 1*occupation + 2*ability + rnorm(10000)
)

lm_1 <- lm(wage ~ female, tb)
lm_2 <- lm(wage ~ female + occupation, tb)
lm_3 <- lm(wage ~ female + occupation + ability, tb)

stargazer(lm_1,lm_2,lm_3,
           type = "html",
           column.labels = c("Biased Unconditional", "Biased", "Unbiased Conditional")
           )
```

# Collider Bias

Dadurch, dass wir im vorherigen Beispiel auf den Beruf (occupation) kontrolliert haben scheint es so, dass Frauen nicht benachteiligt werden.

Annahmen unserer Simulation:

- + Wir haben die Fähigkeiten (ability) als Zufallsvariable simuliert, welche unabhängig vom Geschlecht ist!
  - + `ability = rnorm(10000)`
- + Wir haben eine Diskriminierung simuliert indem wir für "bessere" Jobs höhere Fähigkeiten verlangen, aber weniger Frauen zulassen
  - + `occupation = 1 + 2*ability + 0*female - 2*discrimination + rnorm(10000)`
- + Wir haben das Gehalt so simuliert, dass es für Frauen niedriger ist (hier als `discrimination` deklariert), aber höher ist, je höher die berufliche Position und die Fähigkeiten
  - + `wage = 1 - 1*discrimination + 1*occupation + 2*ability + rnorm(10000)`

Was ist der tatsächliche direkte Effekt der Diskriminierung auf das Einkommen?

Welche Kontrollvariablen benötigen Sie um den Collider Bias aufzuheben?

# Simulationsstudie zu Collider

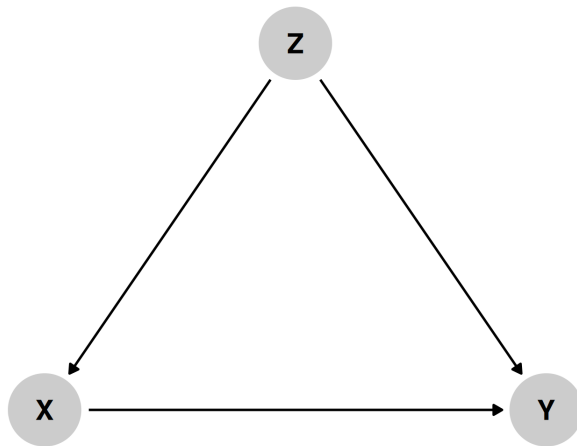
Wir wollen uns die Auswirkungen des Collider Bias näher in einer Simulation anschauen:

	<i>Dependent variable:</i>		
	Biased Unconditional	wage Biased	Unbiased Conditional
	(1)	(2)	(3)
female	-2.970 <sup>***</sup> (0.084)	0.587 <sup>***</sup> (0.029)	-1.014 <sup>***</sup> (0.029)
occupation		1.790 <sup>***</sup> (0.006)	1.007 <sup>***</sup> (0.010)
ability			1.970 <sup>***</sup> (0.022)
Constant	2.014 <sup>***</sup> (0.060)	0.195 <sup>***</sup> (0.020)	1.006 <sup>***</sup> (0.018)
Observations	10,000	10,000	10,000
R <sup>2</sup>	0.110	0.911	0.950
Adjusted R <sup>2</sup>	0.110	0.911	0.950



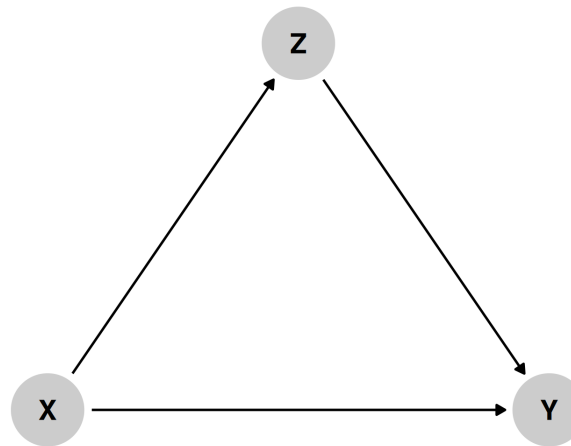
# Welche Arten von Zusammenhang gibt es im DAG?

Confounder



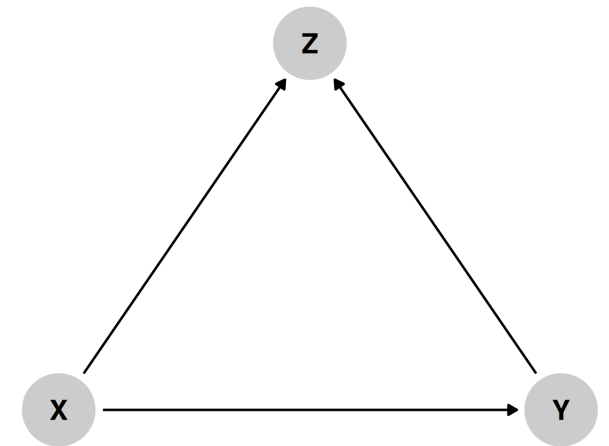
Hier handelt es sich um einen gemeinsamen Ursprung

Mediator



Hier handelt es sich um einen Mediator

Collider



Hier handelt es sich um Selektion / Endogenität