

Directed acyclic graphs (DAG)

Einführung

Ein Ansatz um über Kausalität nachzudenken:

Einführung

Ein Ansatz um 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!

Einführung

Ein Ansatz um über Kausalität nachzudenken:

Directed Acyclic Graphs (DAGs)

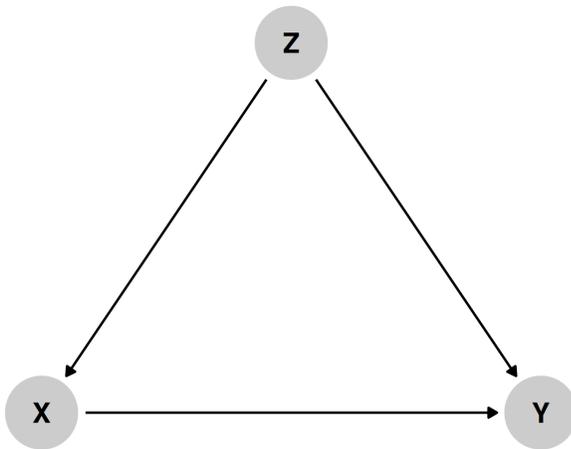
- + Grafische Modelle
- + Kausalität fließt immer in eine Richtung und wird durch Pfeile verdeutlicht
- + Keine Rückwärtskausalität oder Simultanität abbildbar
- + Betrachtet Alternativszenarien
- + Do-Calculus im Hintergrund (Fancy Mathe)

Verwenden wir in dieser Veranstaltung!

DAGs helfen uns insbesondere den zugrunde liegenden Datengenerierungsprozess zu modellieren.

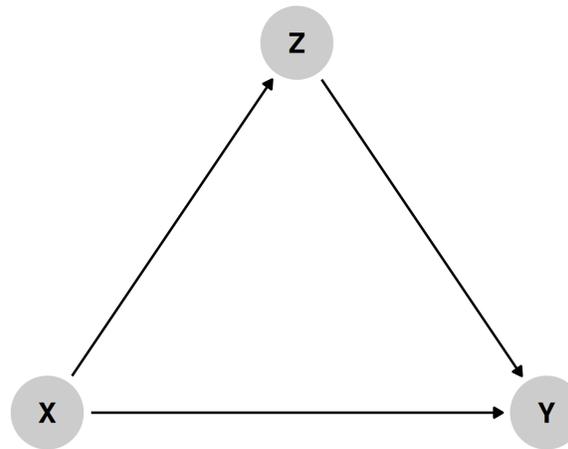
Welche Arten von Zusammenhang gibt es im DAG?

Confounder



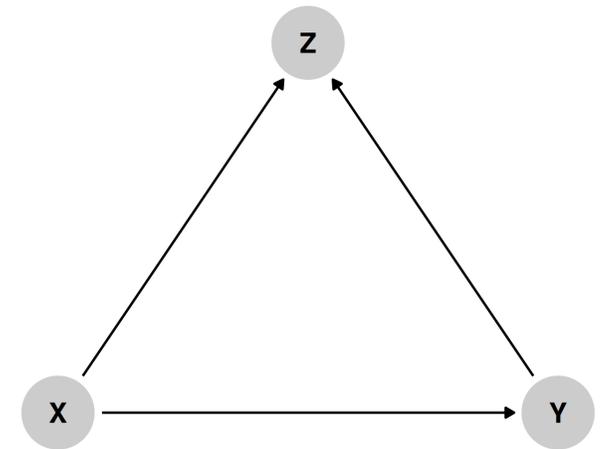
Hier handelt es sich um einen gemeinsamen Ursprung

Mediator



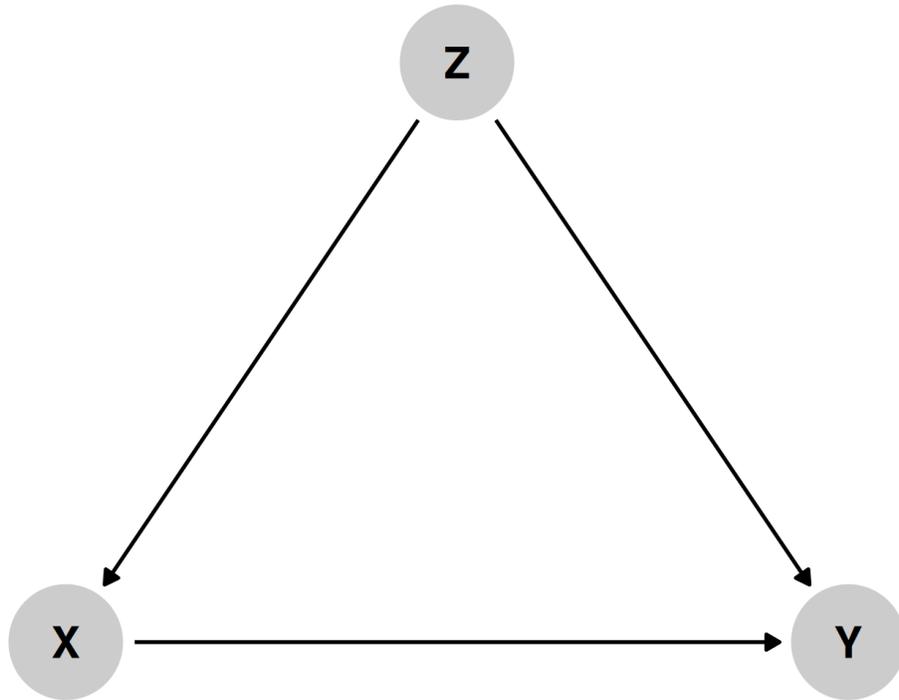
Hier handelt es sich um einen Mittler des Gesamteffekts

Collider



Hier handelt es sich um Selektion / Endogenität

Confounder



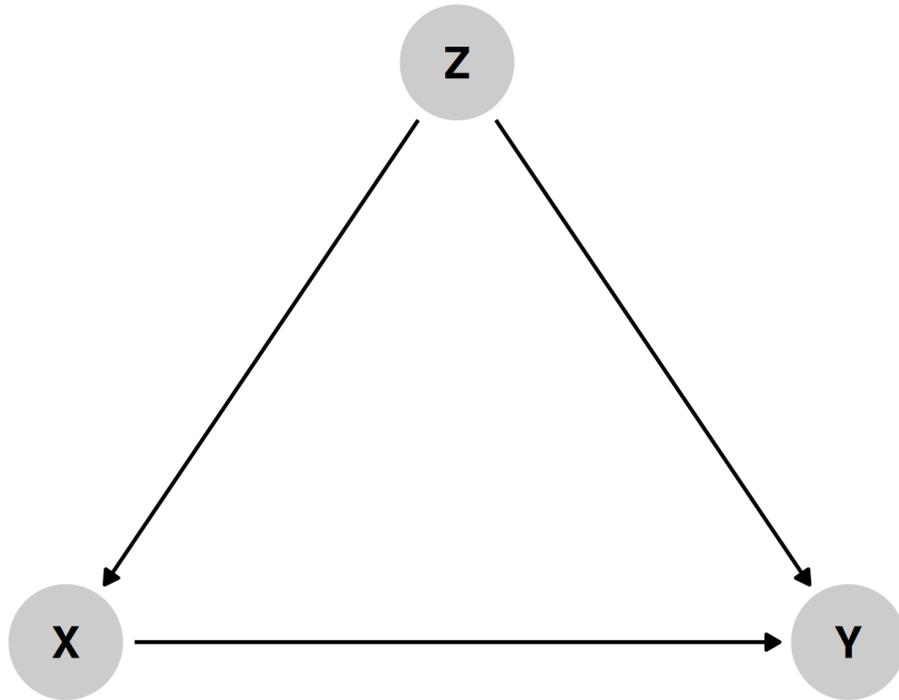
X führt zu Y

+ 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

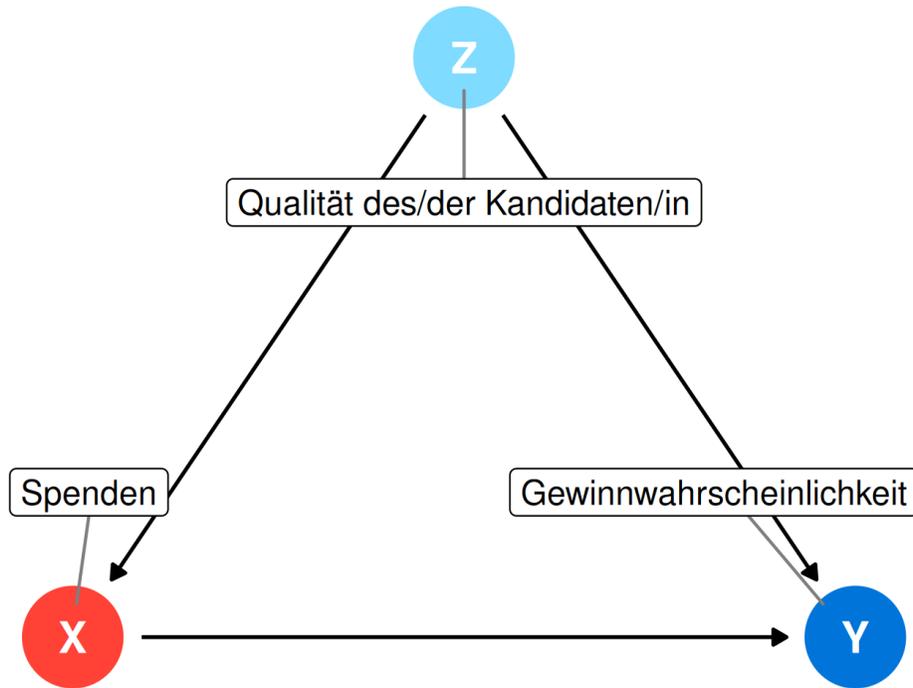
+ Dies sehen wir am *kausalen* Pfad $X \rightarrow Y$

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



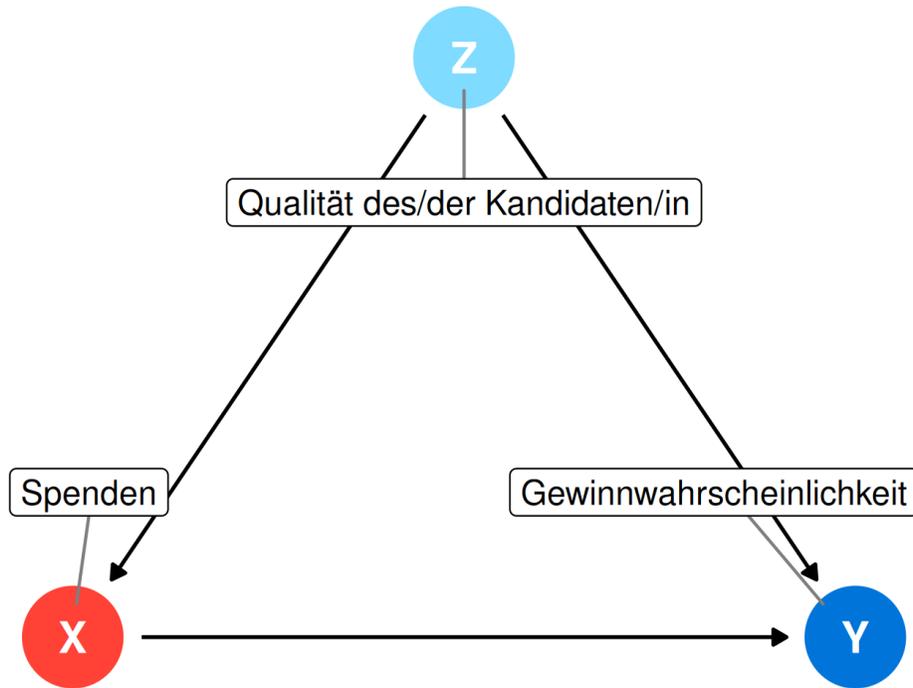
Pfade zwischen **Spenden** und **Gewinnwahrscheinlichkeit**

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**

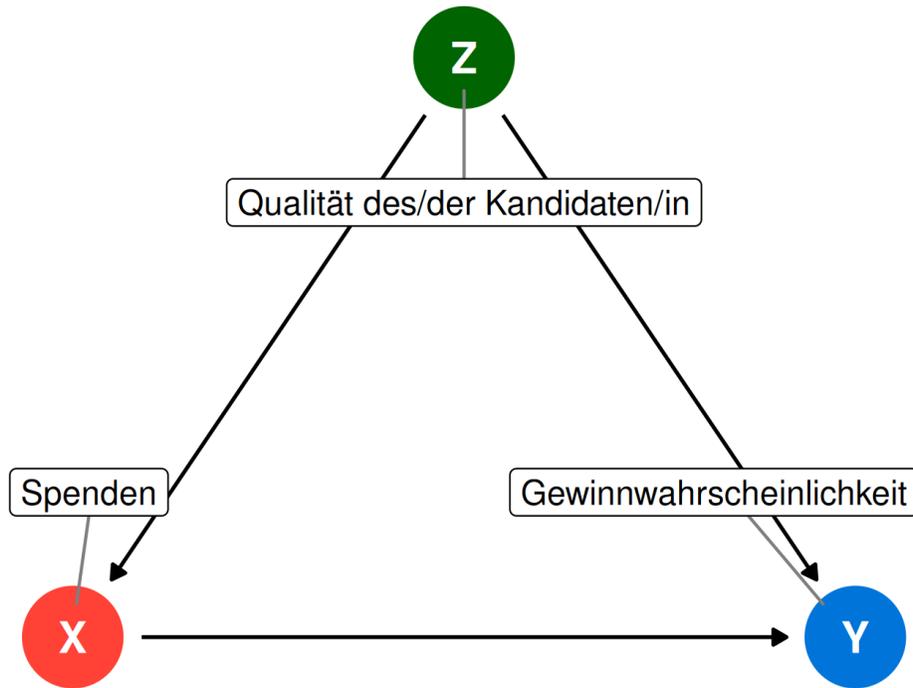
Spenden → Gewinnwahrscheinlichkeit

Spenden ← Qualität des/der Kandidaten/in → Gewinnwahrscheinlichkeit

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

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

Türen schließen



Dadurch das Sie auf die Qualität des/der Kandidaten/in kontrollieren:

- + Eliminieren Sie den Effekt der Qualität des/der Kandidaten/in auf die Spenden
- + Eliminieren Sie den Effekt der Qualität des/der Kandidaten/in auf die Gewinnwahrscheinlichkeit

Zusammenhang des verbleibenden Effekts der Spenden auf den verbleibenden Effekt der Gewinnwahrscheinlichkeit ist unser **kausaler Effekt** von Spenden auf Gewinnwahrscheinlichkeit.

→ Durch die Kontrolle vergleichen wir hier Kandidaten, wie wenn diese die gleiche Qualität hätten

Kontrollieren innerhalb einer Regression

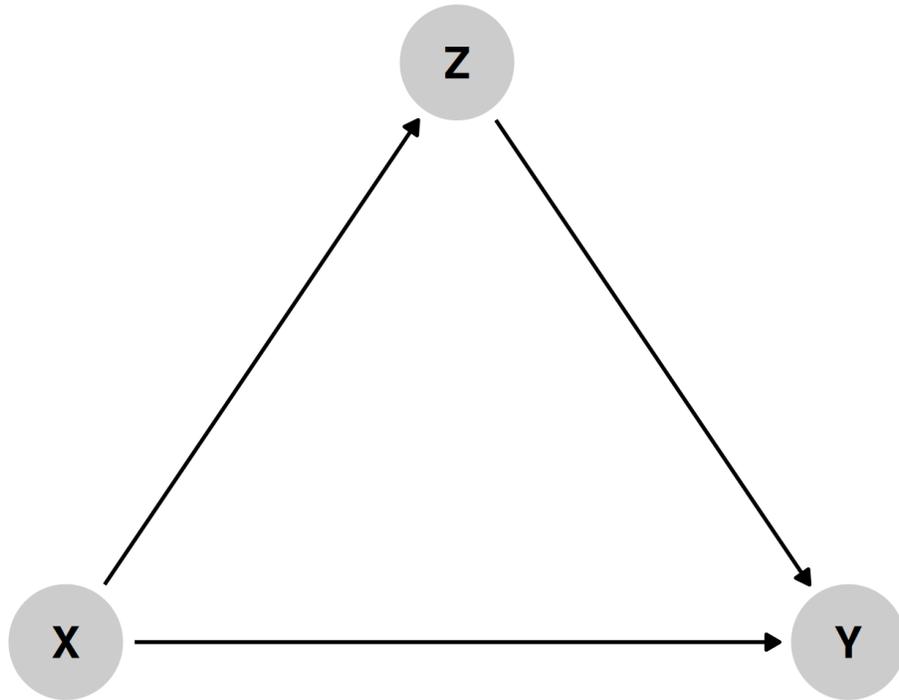
Eine Möglichkeit auf Variablen zu "kontrollieren" ist mittels einer multiplen linearen Regression:

$$\text{Gewinnwahrscheinlichkeit} = \beta_0 + \beta_1 * \text{Spenden} + \beta_2 * \text{Qualität} + \epsilon$$

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

- + Matching
- + Stratifizierung
- + Synthetische Kontrollgruppen

Mediator

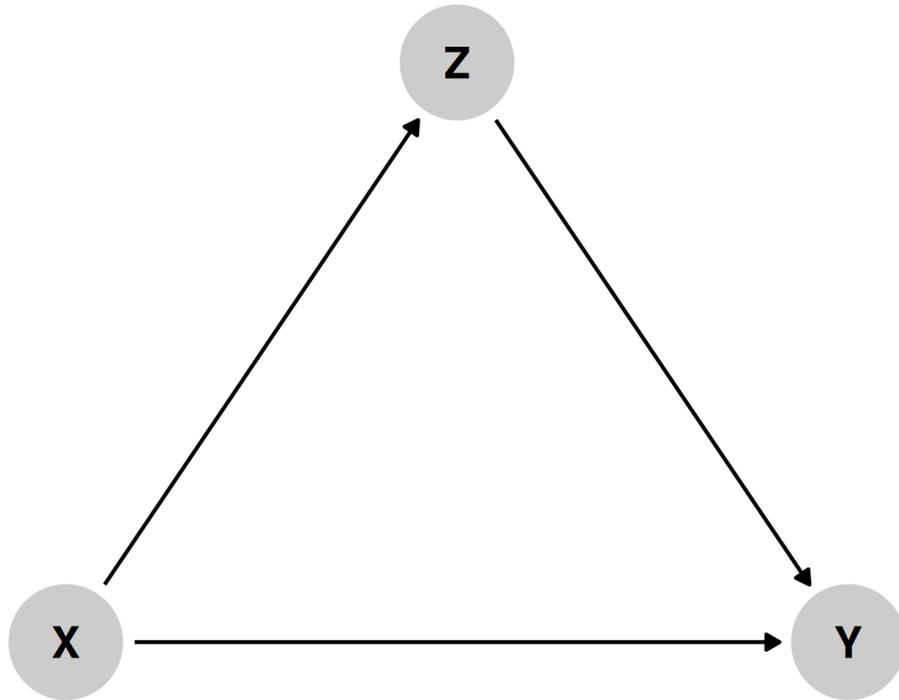


X führt zu Y

X führt zu Z, welches wiederum zu Y führt

Sollten wir für Z kontrollieren?

Mediator



X führt zu Y

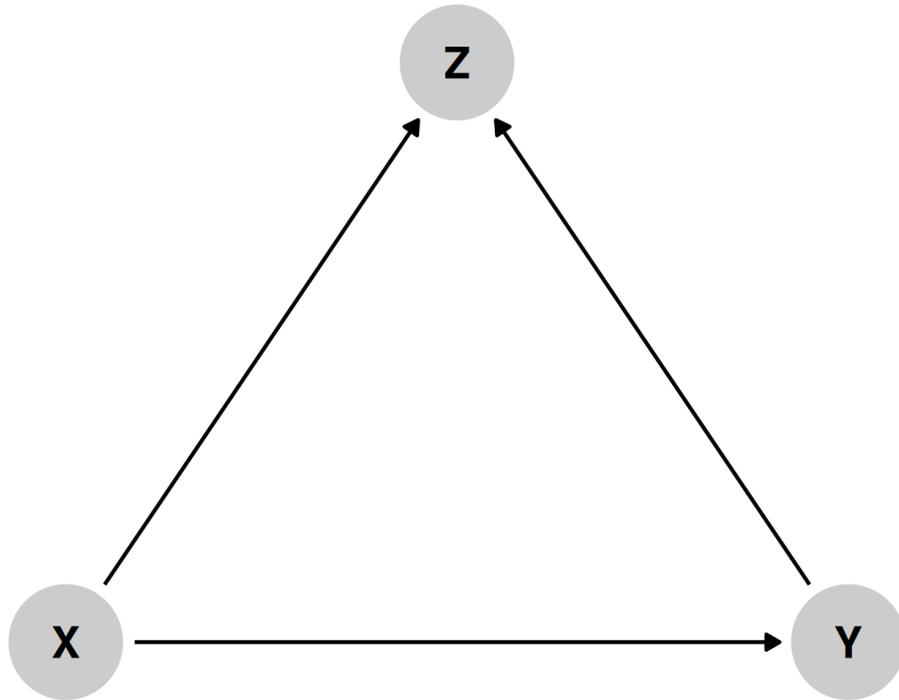
X führt zu Z, welches wiederum zu Y führt

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.

Collider



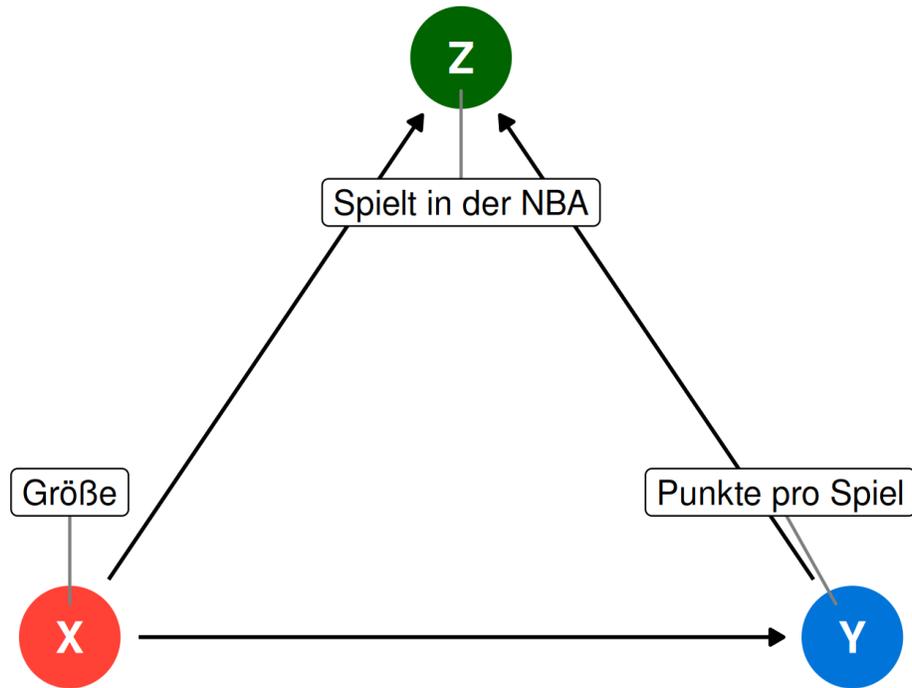
X führt zu Y

X führt zu Z

Y führt zu Z

Sollten wir für Z kontrollieren?

Beispiel zu Collidern

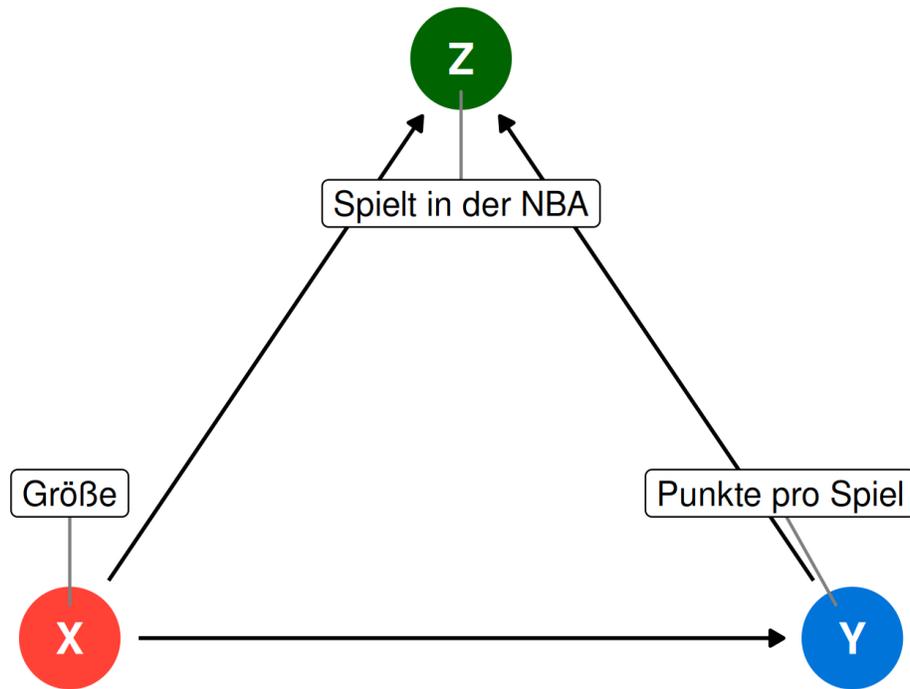


Größe \rightarrow Punkte pro Spiel

Größe \rightarrow Spielt in der NBA \leftarrow Punkte pro Spiel

Warum ist die Variable "Spielt in der NBA" ein Collider?

Beispiel zu Collidern



Größe → Punkte pro Spiel

Größe → Spielt in der NBA ← Punkte pro Spiel

Warum ist die Variable "Spielt in der NBA" ein Collider?

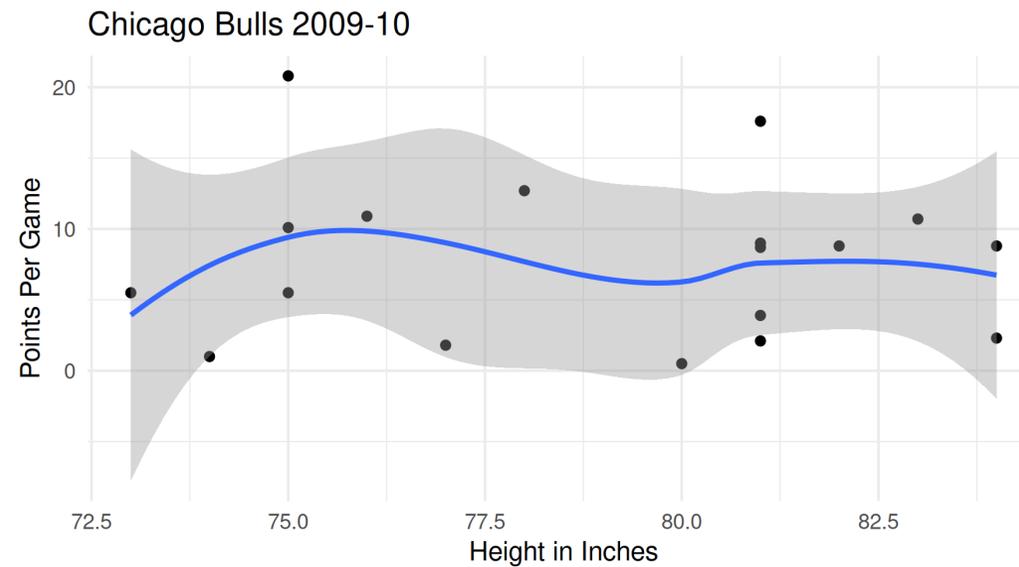
Wenn wir auf darauf kontrollieren, dass die Person in der NBA spielt, öffnen wir den zweiten Kanal, da "Spielt in der NBA" als Collider fungiert.

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

Beispiel zu Collidern

In der Gesamtbevölkerung gibt es vermutlich schon einen Zusammenhang zwischen der Körpergröße und den Punkten pro Spiel im Basketball.

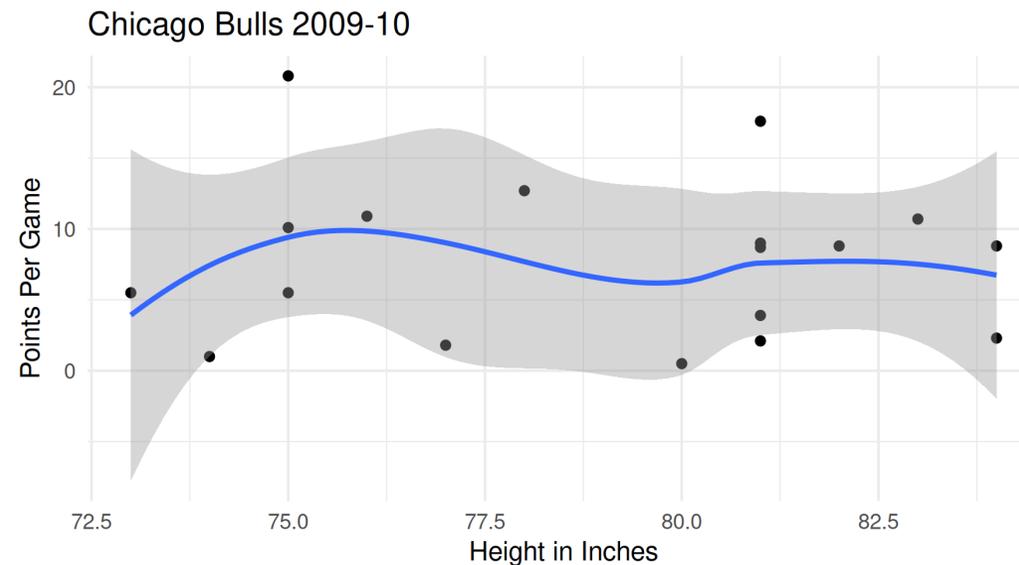
Im Datensatz betrachten wir jedoch nur Spieler, die in der NBA spielen und damit sowohl groß sind, als auch gut Basketball spielen können!



Beispiel zu Collidern

In der Gesamtbevölkerung gibt es vermutlich schon einen Zusammenhang zwischen der Körpergröße und den Punkten pro Spiel im Basketball.

Im Datensatz betrachten wir jedoch nur Spieler, die in der NBA spielen und damit sowohl groß sind, als auch gut Basketball spielen können!



- + Collider können tatsächliche kausale Effekte verdecken
- + Collider können scheinbar kausale Effekte erzeugen

Überlegungen zu den DAGs

"Big data" hilft nicht um kausale Fragestellungen zu beantworten!

- + Um kausale Fragestellungen beantworten zu können müssen wir den Datengenerierungsprozess verstehen
 - + Woher stammen unsere Daten und welche Effekte hat es auf bestimmte Variablen zu kontrollieren?
- + Wir müssen die institutionellen Gegebenheiten kennen um glaubwürdige Identifikationsstrategien entwickeln zu können

DAGs können uns hier sehr gute Dienste erweisen.