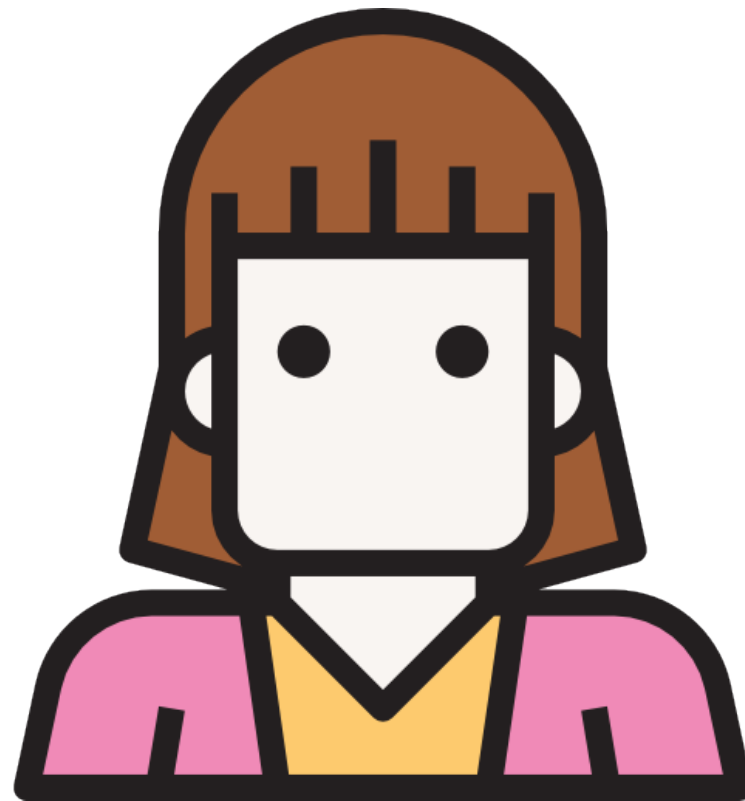


Warum Statistik nichts hilft zur Entscheidungsfindung

Einführung in die Kausalanalyse

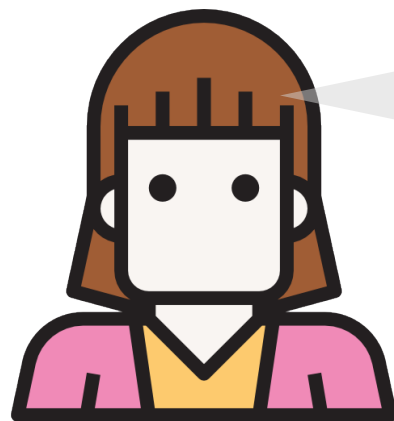
Angi, Don und Wolfi

Fallstudie: Angela hat einen neuen Job, als Marktforscherin



Angela M.,
Marktforscherin

Frisch nach Saratoga County gezogen, schicke Gegend



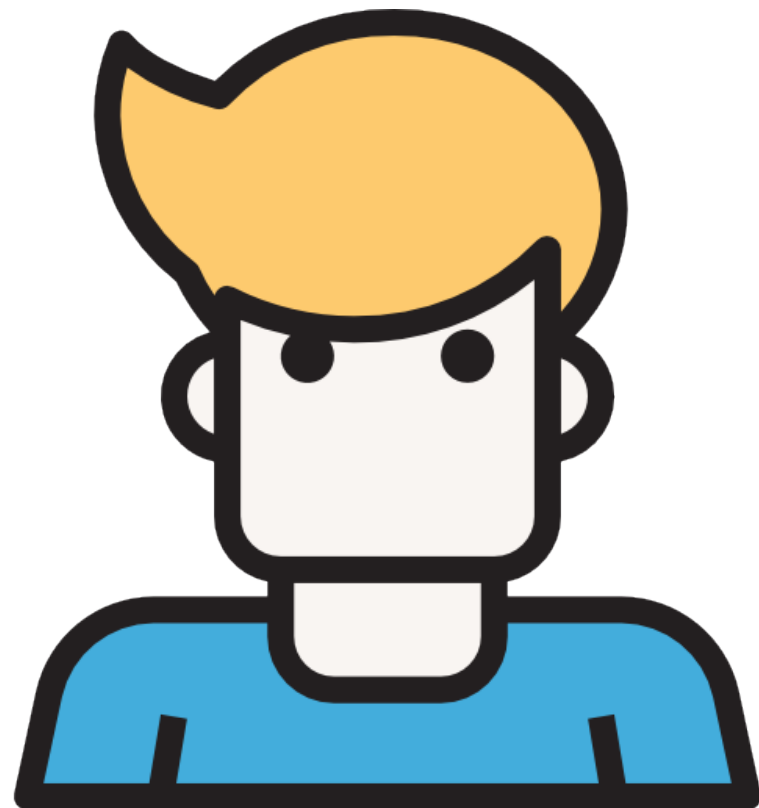
Yeah, I love my new job!

Angi



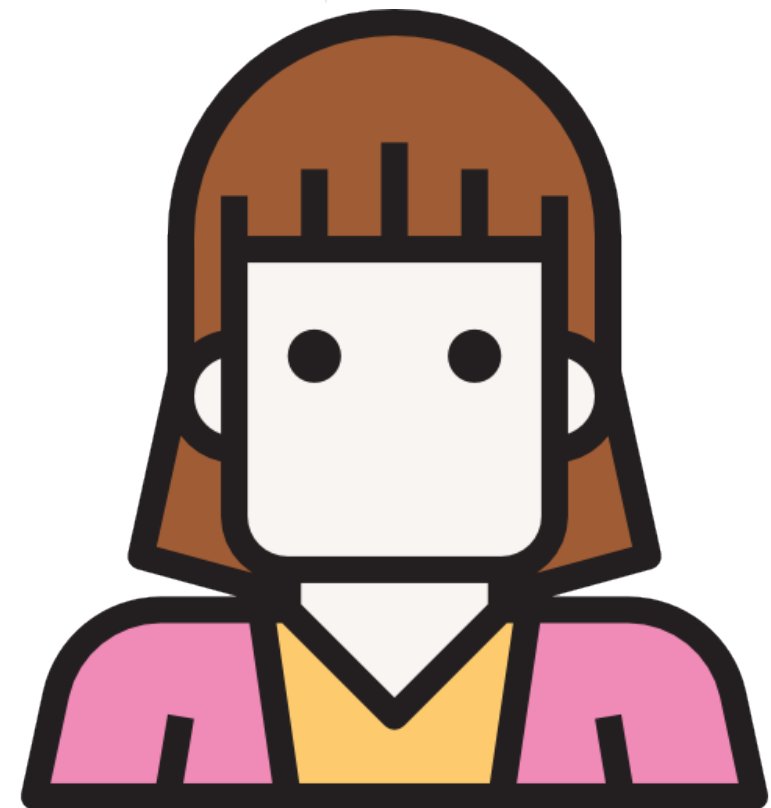
Neuer Job, erste Aufgabe: Verkaufswert von Immobilie vorhersagen

Wieviel ist mein Haus wert!
So groß, tolles Haus!



Don, Immobiliengigant

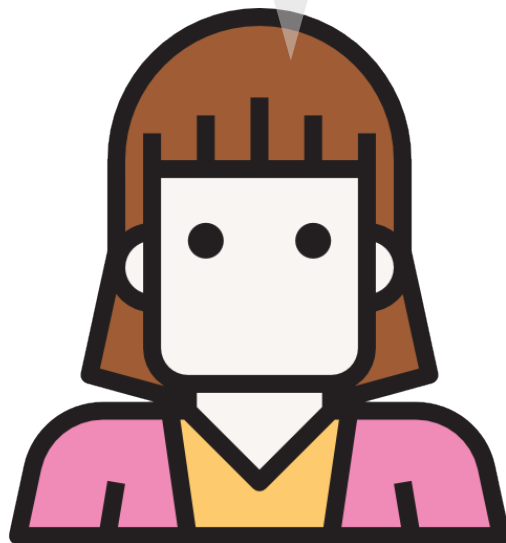
Find ich heraus.
Ich mach's
wissenschaftlich.



Angi

Angi schaut sich Immobilien-Daten zum Landkreis an

Ich liebe Daten! 🥰

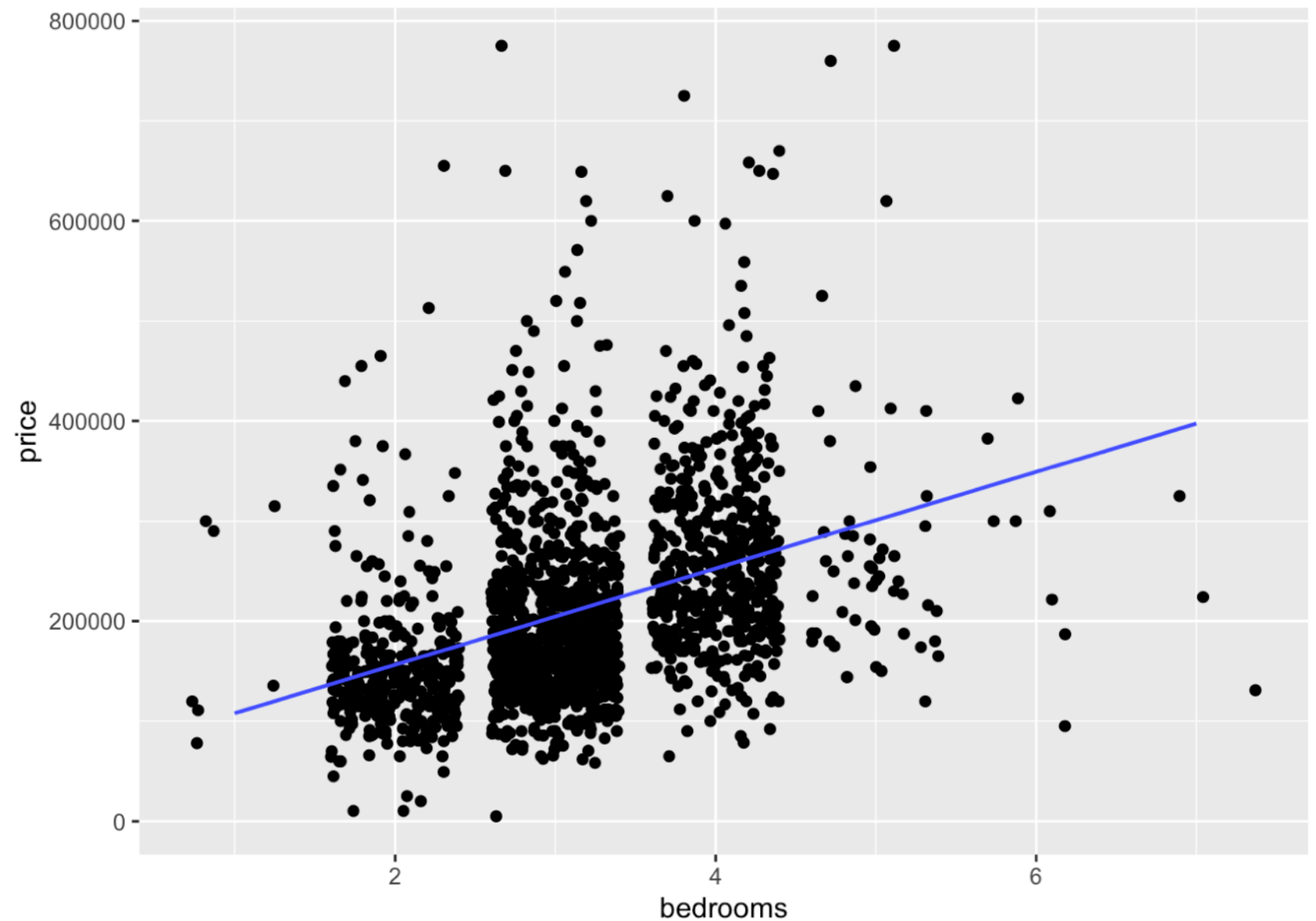
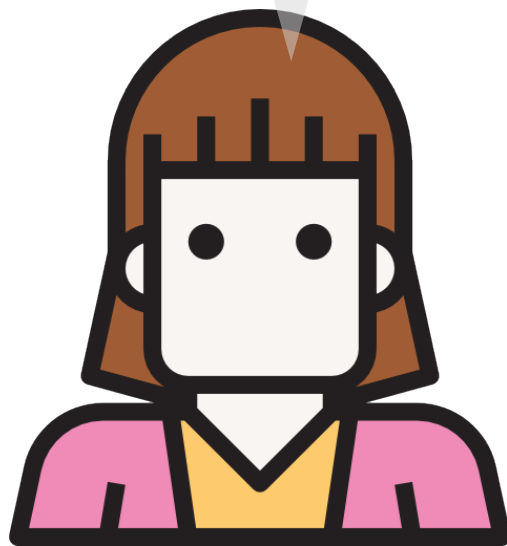


Angi

price	square-feet	age	a/c	fireplace	heating	...
132500	84.17	42	No	Yes	Electricity	
181115	181.44	0	No	No	Gas	
109000	180.60	133	No	Yes	Gas	
155000	180.60	13	No	Yes	Gas	
86060	78.04	0	Yes	No	Gas	
120000	107.02	31	No	Yes	Gas	
153000	255.67	33	No	Yes	Oil	
170000	154.40	23	No	Yes	Oil	
90000	151.62	36	No	No	Electricity	
122900	131.55	4	No	No	Gas	
...	

Model 1: Preis als Funktion der Anzahl der Zimmer

Hey Don!
Mehr Zimmer,
mehr Kohle!



Angi berichtet Don ihre Ergebnisse

```
library(mosaic)
```

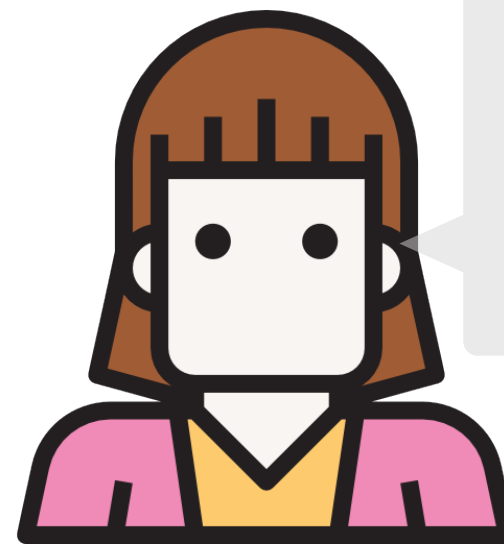
```
data("SaratogaHouses")
```

```
modell <- lm(price ~ bedrooms,  
data = SaratogaHouses)  
coef(modell)
```

```
## (Intercept)      bedrooms  
##      59862.96      48217.81
```

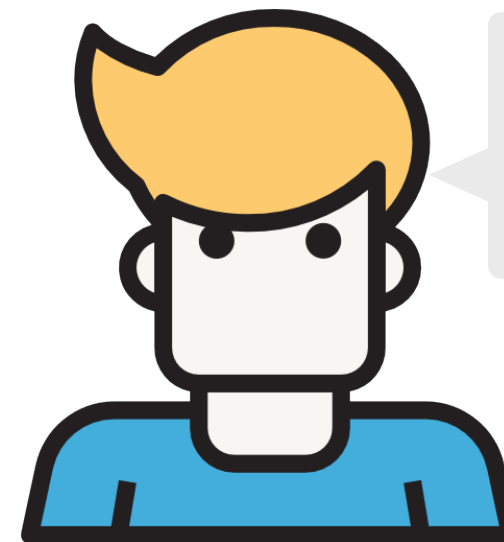
```
dons_house <- data.frame(bedrooms = 2)  
predict(modell, dons_house)
```

```
##           1  
## 156298.6
```



Angi

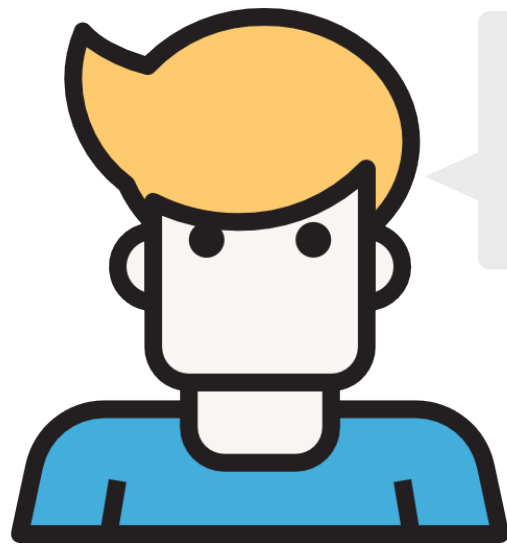
Jedes Zimmer ist knapp 50 Tausend wert. Dein Haus hat einen Wert von 150 Tausend.



Don

Zu wenig! 🤔 😡 🤯

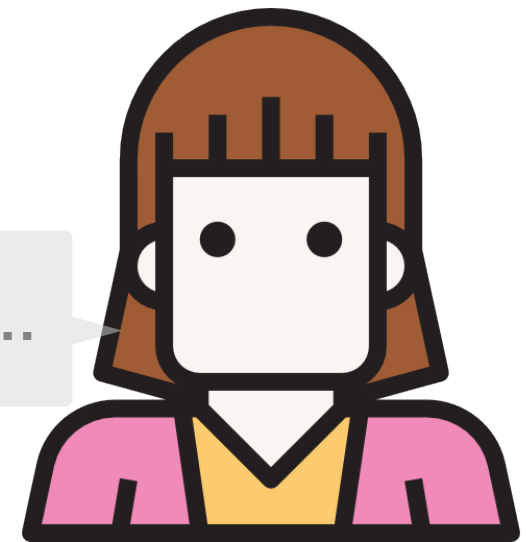
Don hat eine Idee: Jedes Zimmer in zwei Teile teilen!



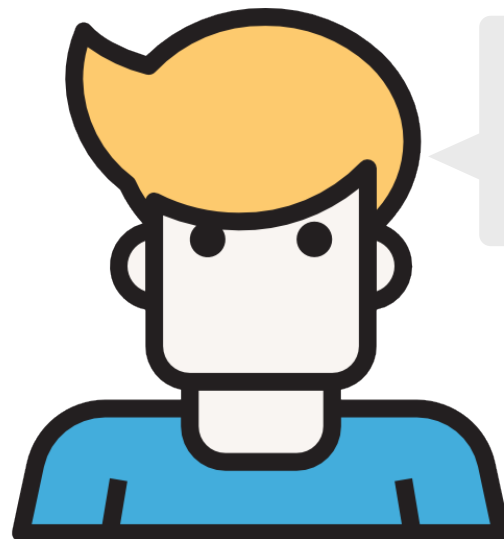
Don

Ich bau eine Mauer!
In jedes Zimmer!

Moment mal ...



Angi



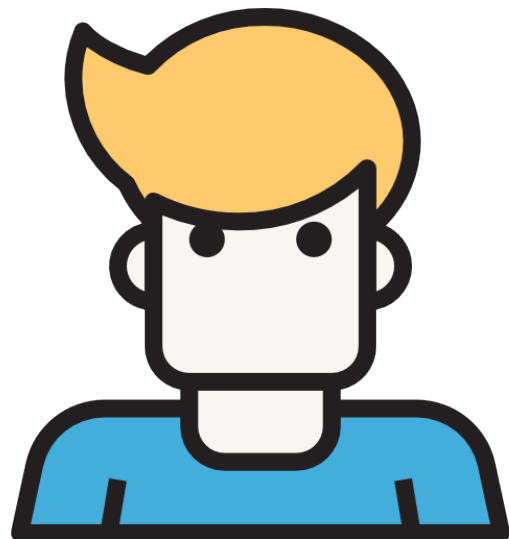
Rechne meinen Preis aus!
Jetzt!

Mit 4 (statt 2) Schlafzimmer steigt der Wert auf 250 Tausend

Laut Modell 1

```
dons_new_house <- data.frame(bedrooms = 4)  
predict(modell1, dons_new_house)
```

```
##           1  
## 252734.2
```

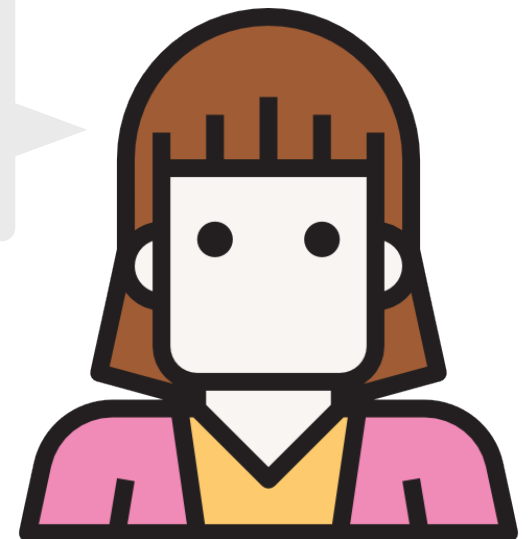


Don

Volltreffer! Jetzt verdiene ich 100 Tausend mehr!



Moment mal ...



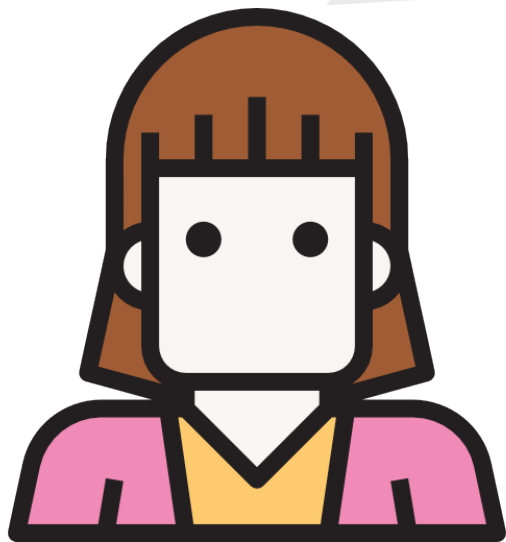
Angi

Modell 2: Preis als Funktion von Zimmerzahl und von Quadratmetern

```
model2 <- lm(price ~ bedrooms + livingArea, data = SaratogaHouses)
coef(model2)
```

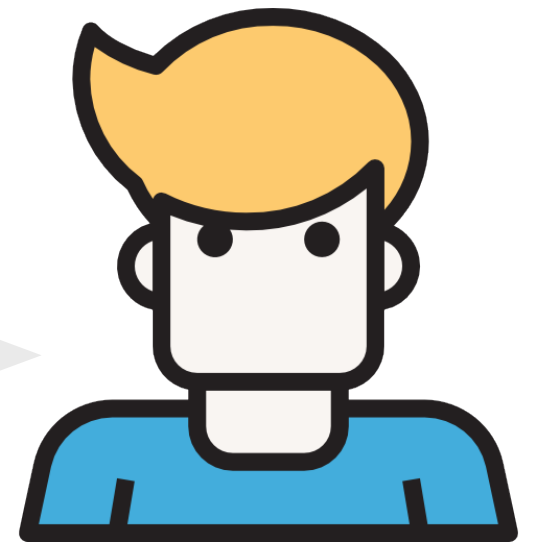
```
## (Intercept)    bedrooms  livingArea
## 36667.895    -14196.769      125.405
```

Die Zimmer halbieren
verringert den Wert,
Don!



Angi

Was, weniger
Geld?! Oh nein!

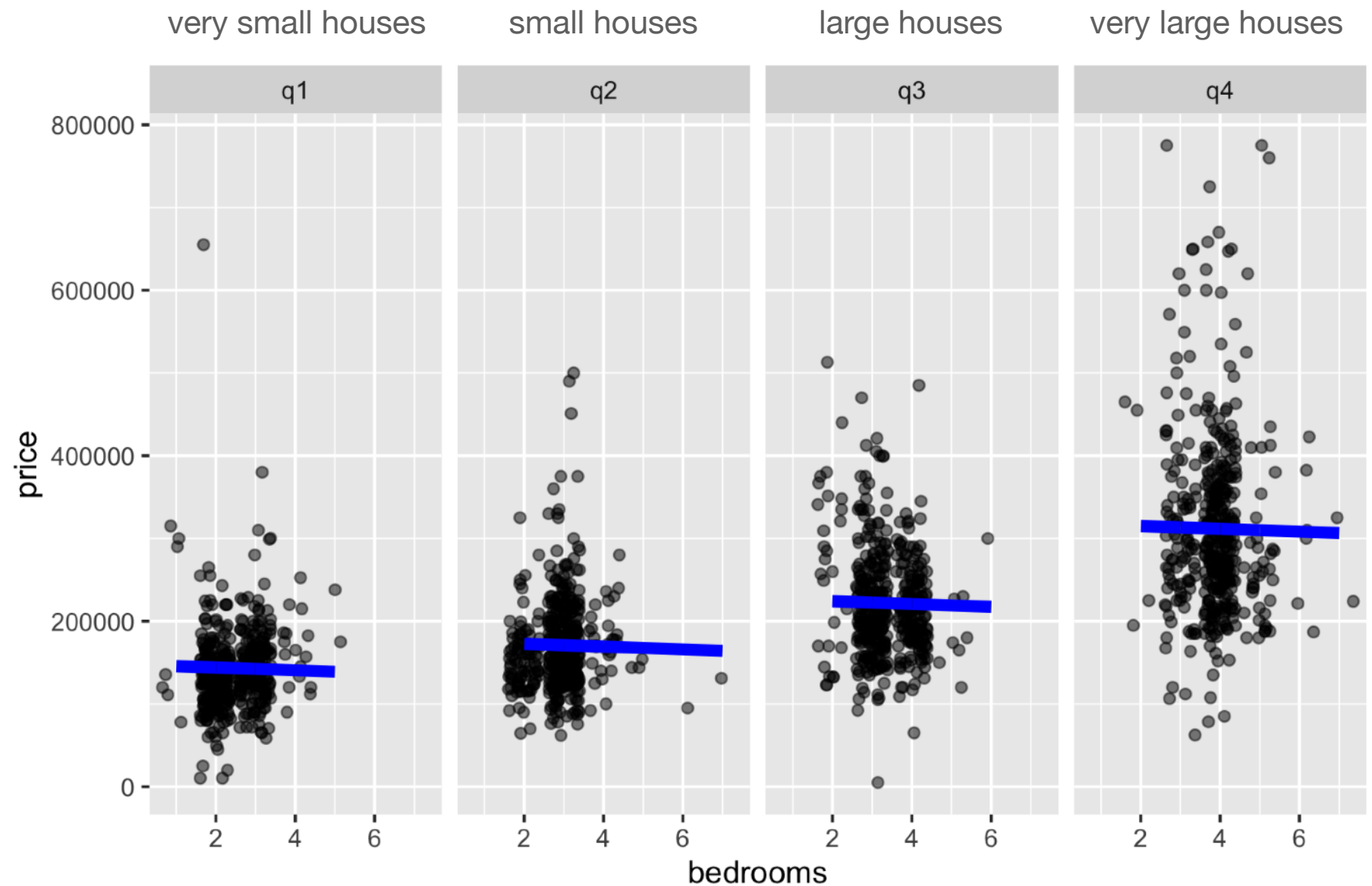
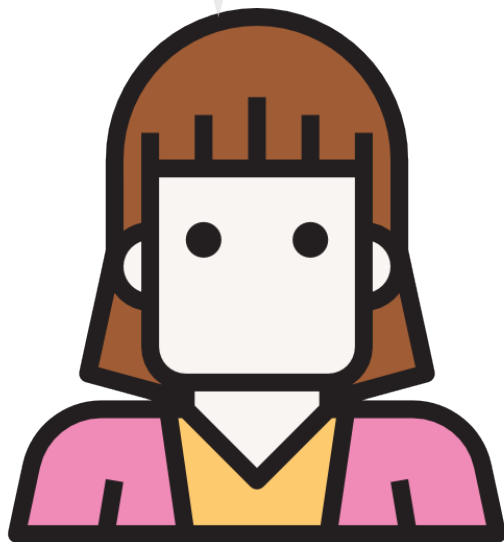


Don

Zimmerzahl ist **NEGATIV** mit dem Preis korreliert

... wenn man die Wohnfläche (Quadratmeter) kontrolliert

NE-GA-TIV!



Take-home message #1

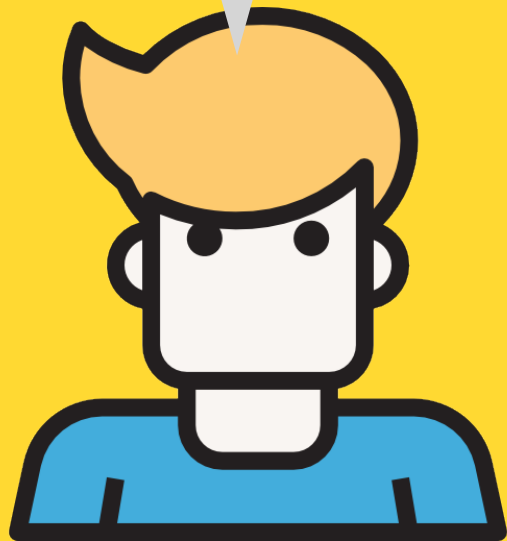
Prädiktoren hinzufügen oder wegnehmen hat oft einen Einfluss auf den Zusammenhang auf die anderen Prädiktoren im Modell.



Wolfi

Take-home message #2

Aber welchem Modell soll ich jetzt glauben?



Don

Keine Antwort von Statistik.



Wolfi

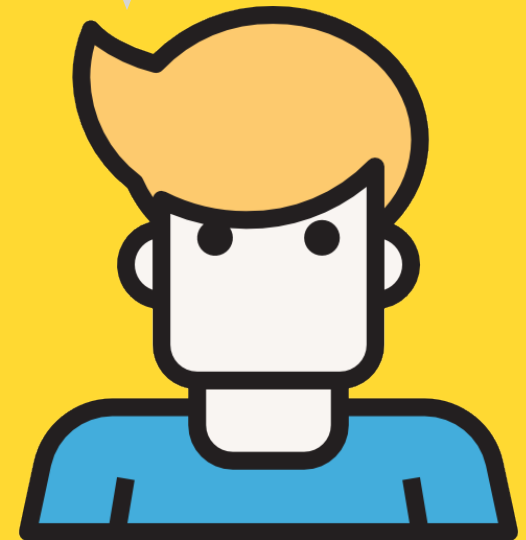
Take-home message #3

Beobachtungsstudien taugen oft nicht, um Entscheidungen zu treffen.



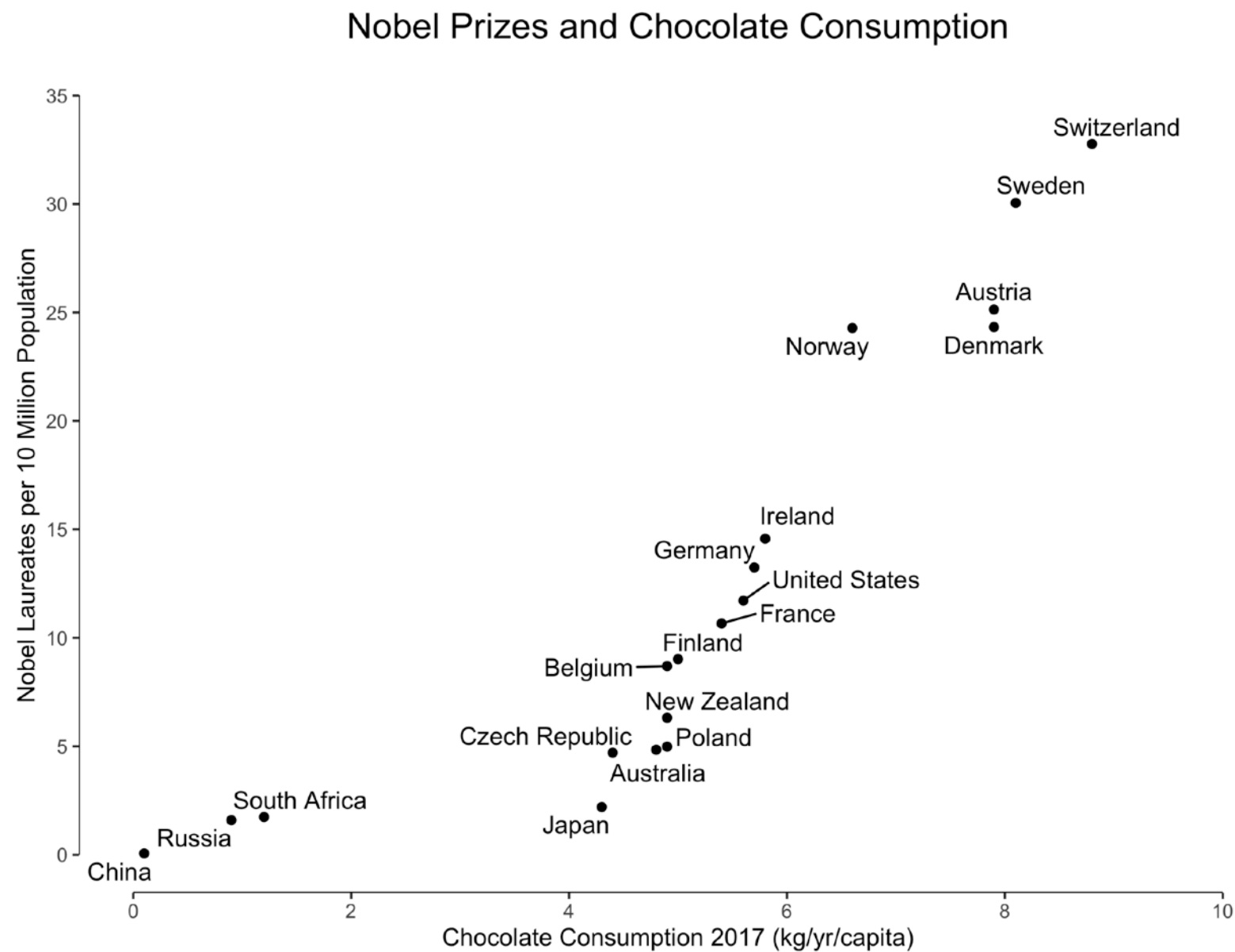
Wolfi

Hast du was gegen Wissenschaft?

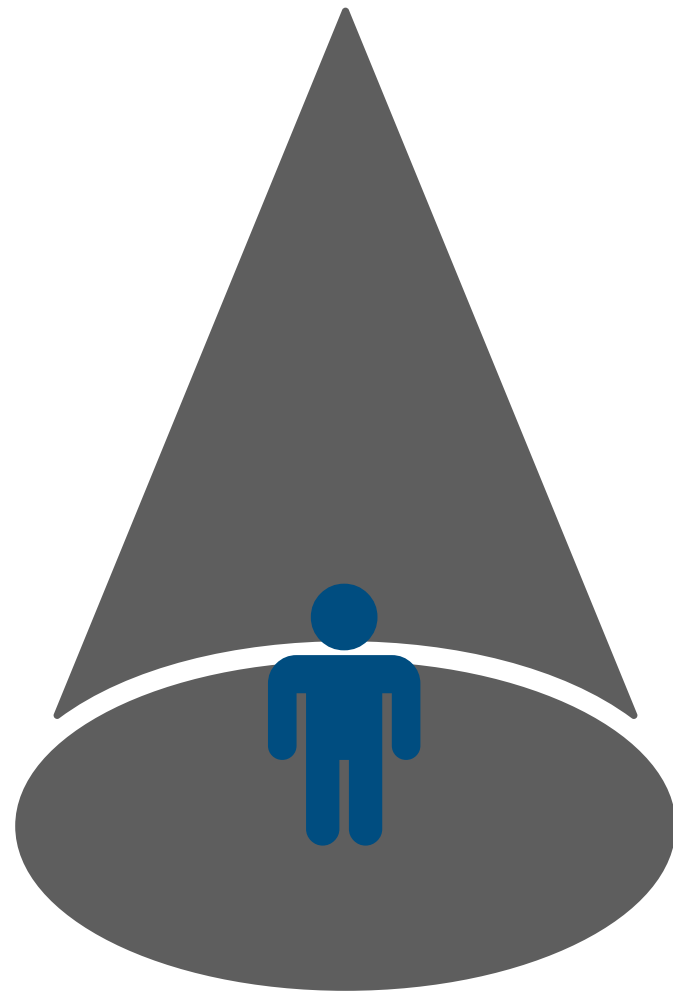


Don

Schoko futtern macht Nobelpreise?



Statistische Zusammenhänge sind echt ... oder trügen

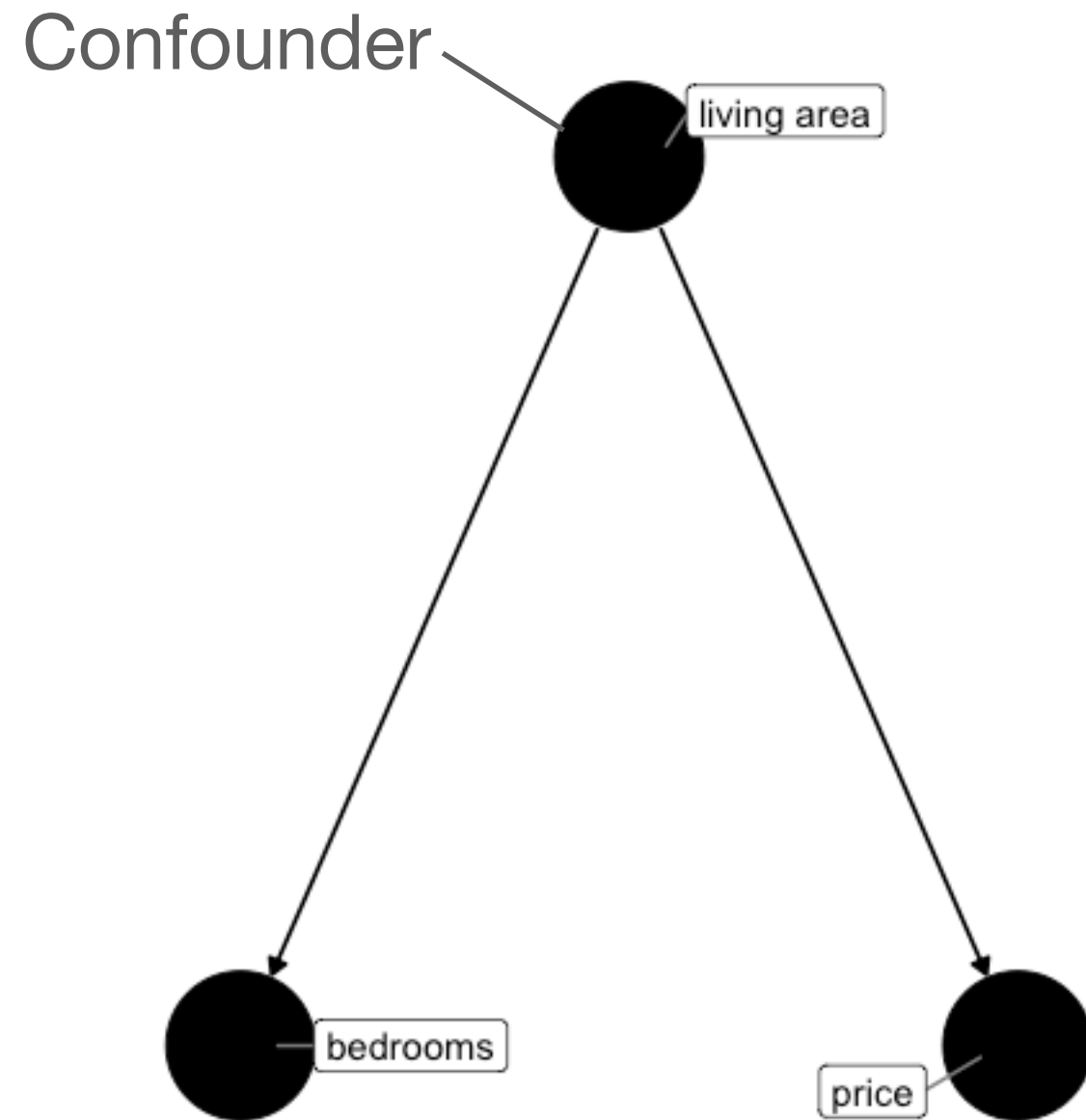


falsch: Scheinzusammenhang

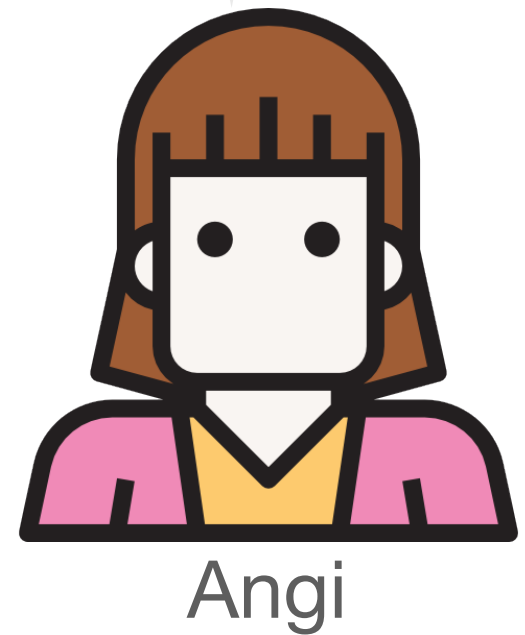


wahr: Kausalzusammenhang

Model 2: Wohnfläche als gemeinsame Ursache (Konfundierung)

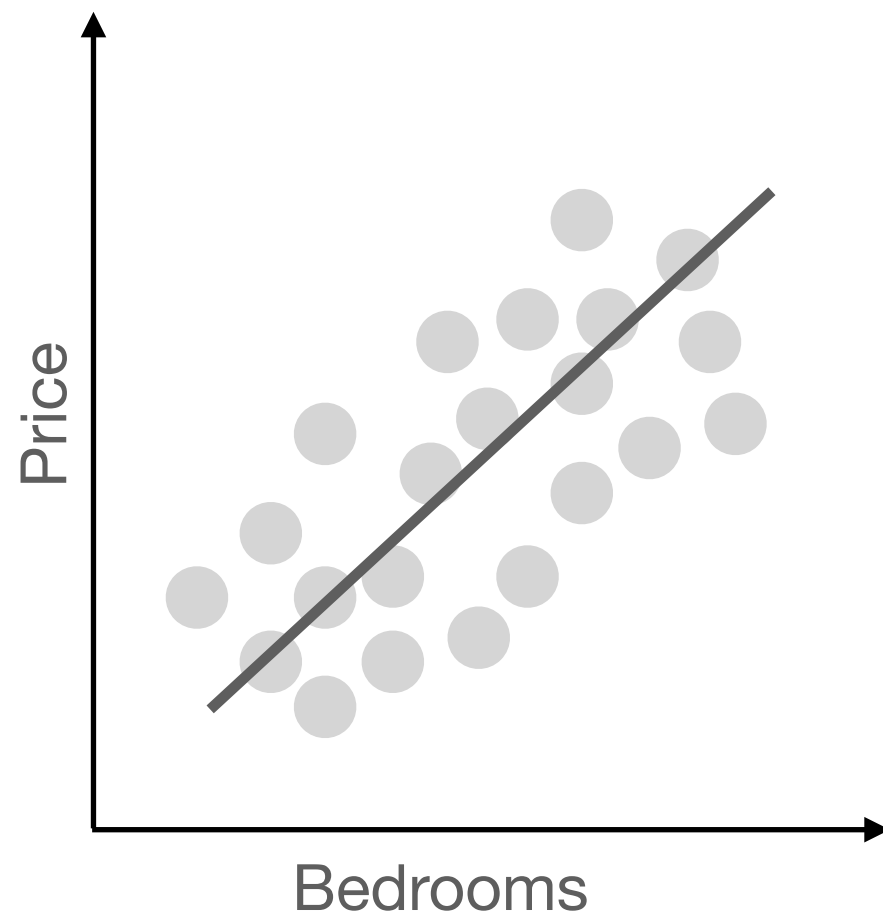


Model 2: Wir
adjustieren den
Preis.



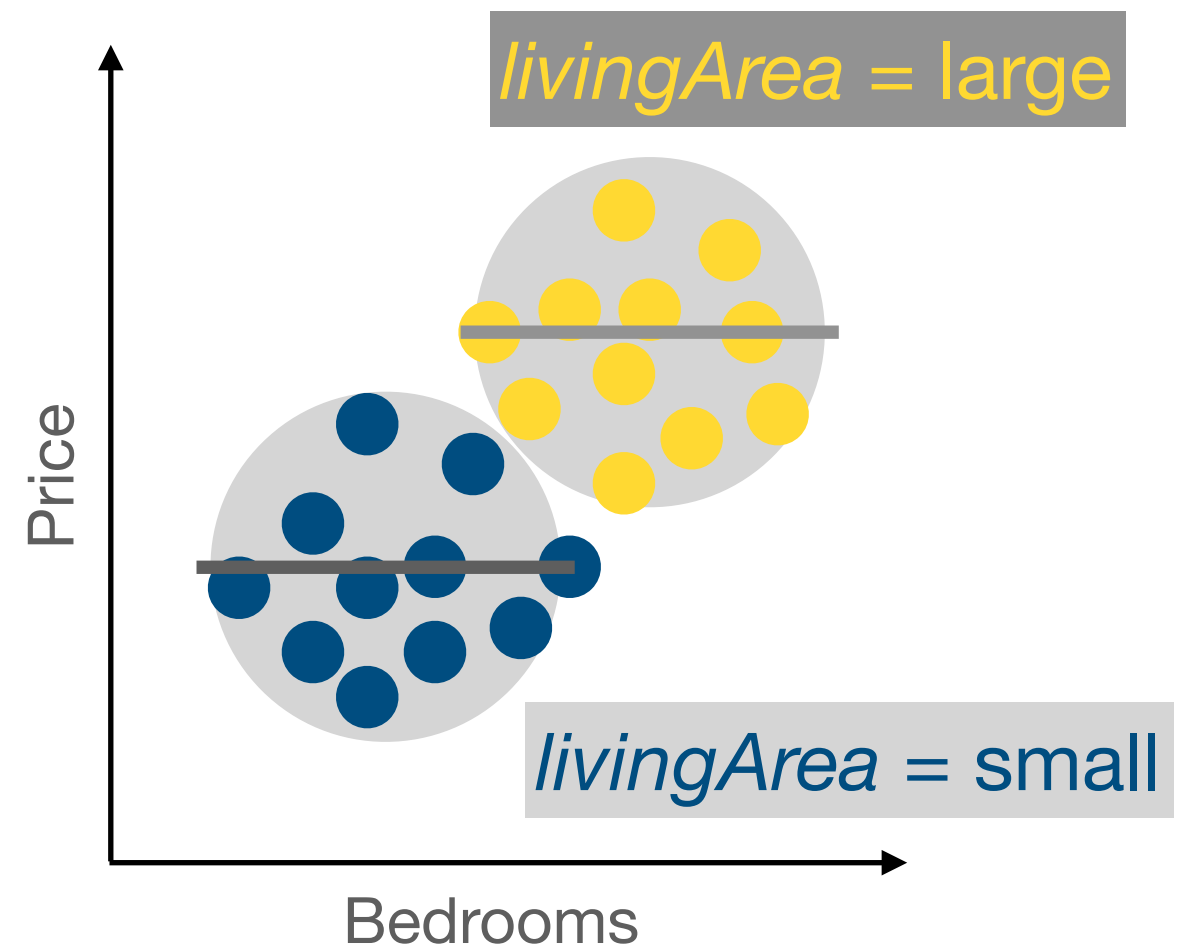
Konfundierer muss man kontrollieren

Model 1: Konfundierer *Wohnfläche*
nicht kontrolliert



Scheinkorrelation tritt auf

Model 1: Konfundierer *Wohnfläche*
ist kontrolliert



Scheinkorrelation verschwindet

Take-home message #4

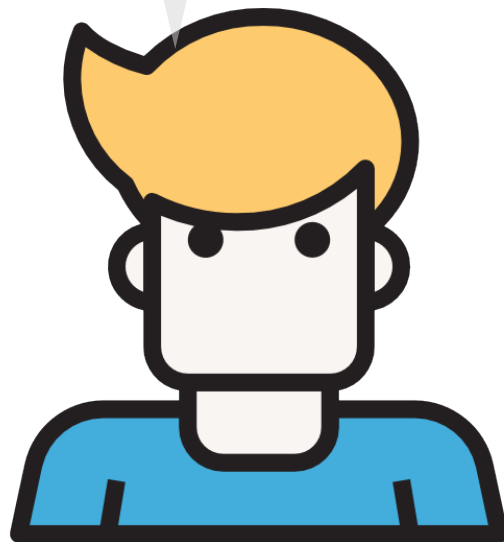


Wolfi

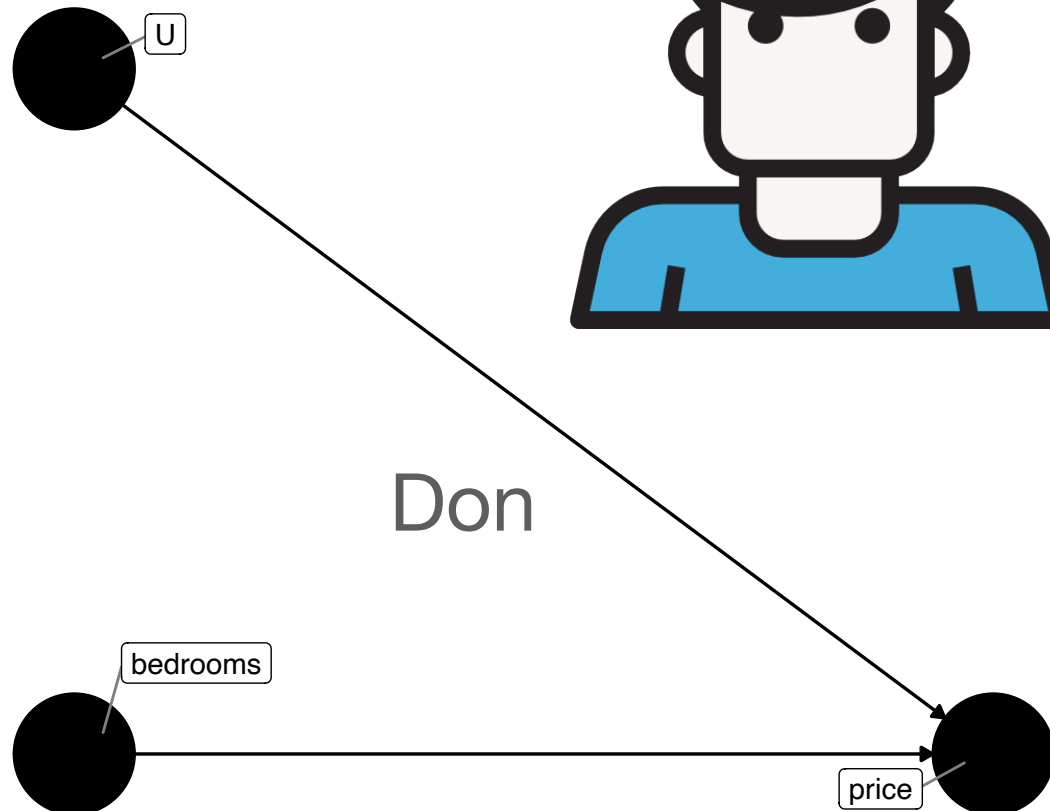
Nur mit einem Kausalmodell kann man
Scheinkorrelationen von echten, kausalen
Zusammenhängen unterscheiden

Modell 1 passt nicht zu den Daten

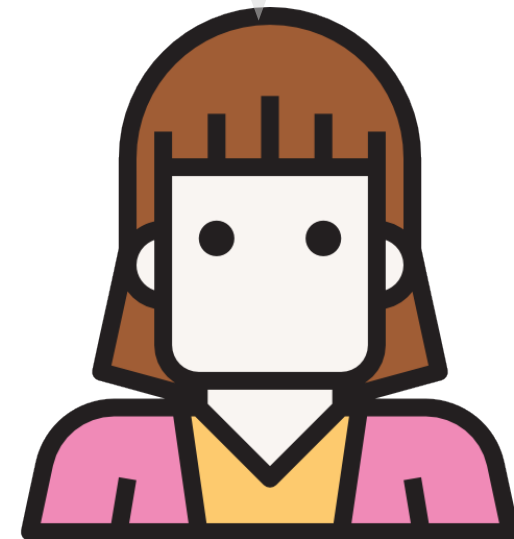
Modell 1 sagt, es gibt keine Konfundierung!



Don



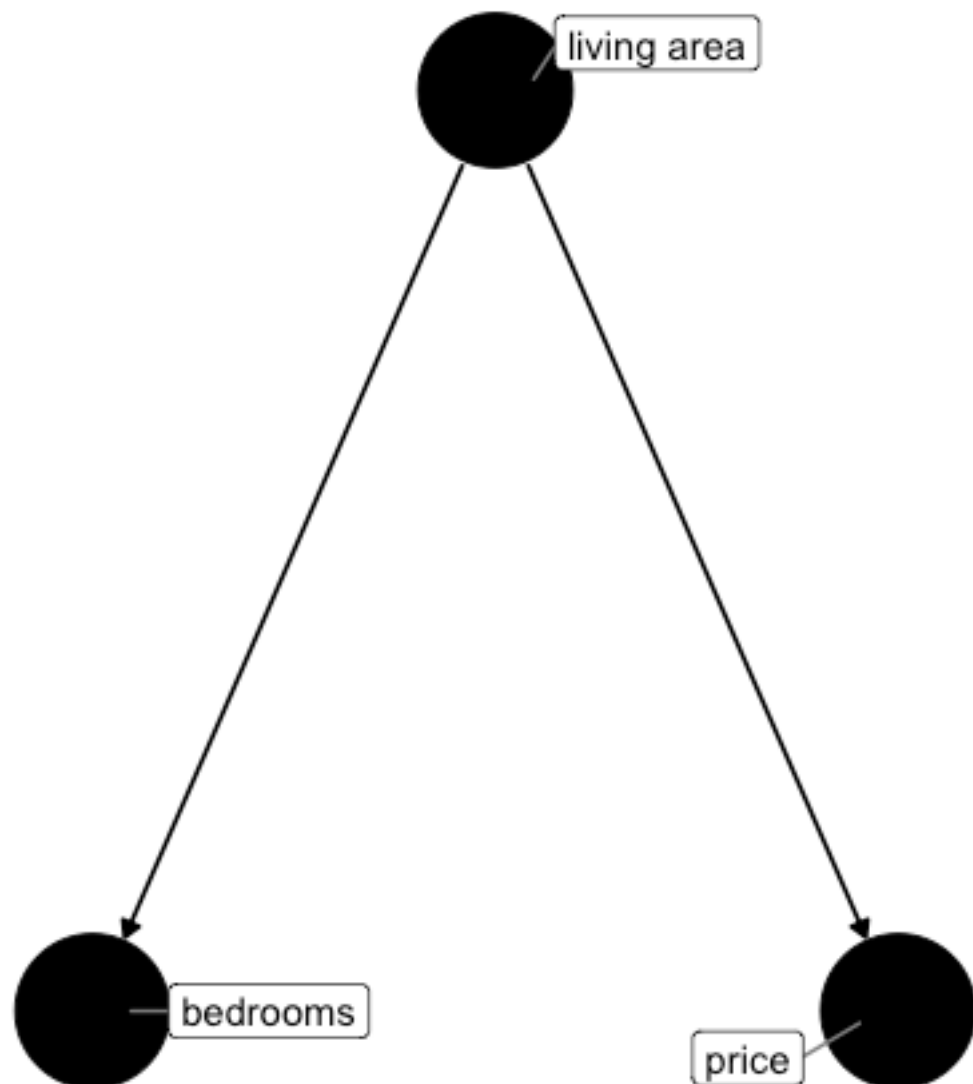
Modell 2 zeigt, dass es Konfundierung gibt. Dein Modell ist falsch, Don.



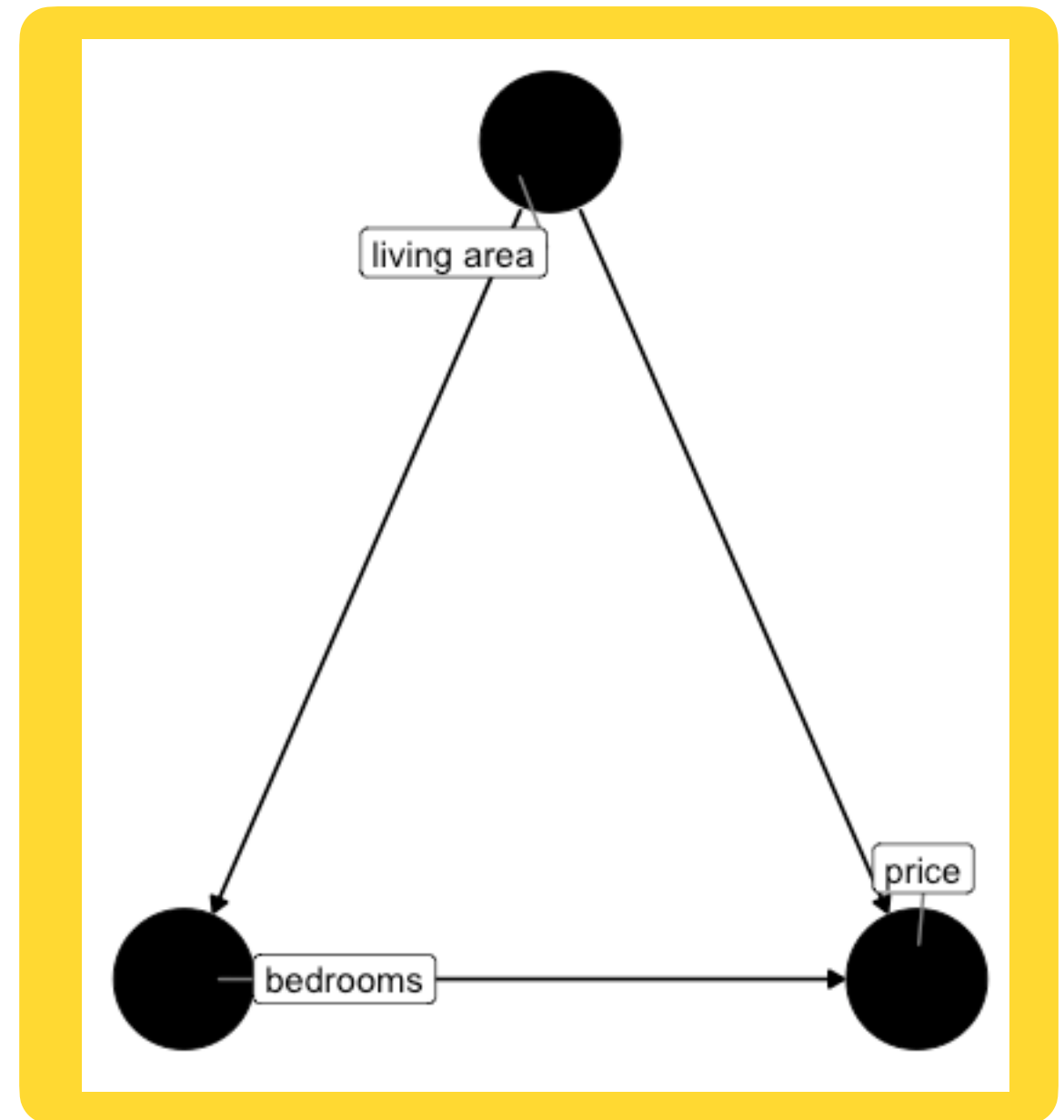
Angi

Angis Modell ist aber auch nicht richtig

Angis Modell (Modell 2)



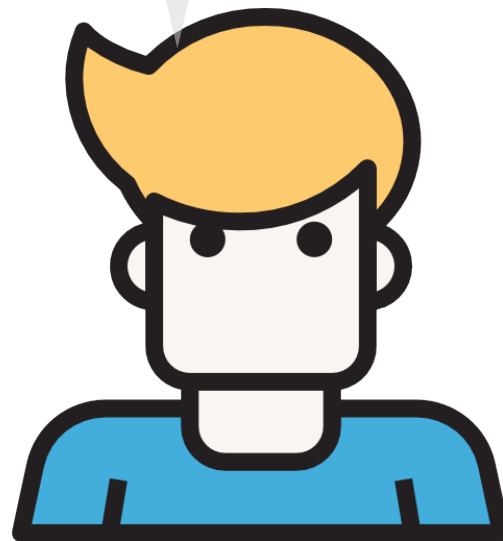
Wolfis model (Modell 3)



Wolffs Modell passt am besten auf die Daten.

Überlegen Sie sich ein Kausalmodell für Dons Problem

Schaffst du nie,
du bist gefeuert!



jklöj

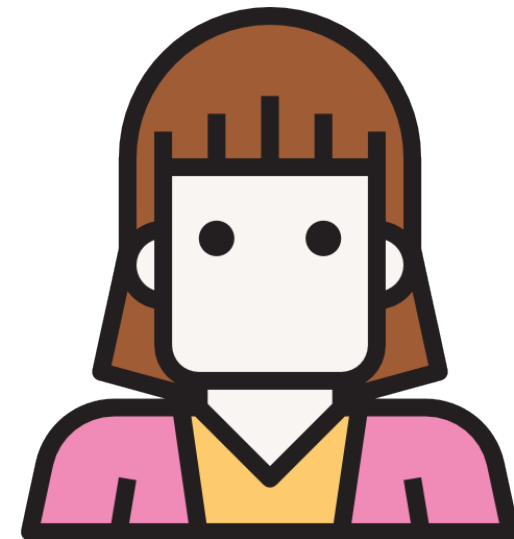
Es gibt drei Arten von wissenschaftlichen Studien, sagt Wolfi

Das Gute, das
Böse und das
Hässliche



Wolfi

Wolfi, reiß dich
bitte zusammen!



Angi

Es gibt drei Arten von wissenschaftlichen Studien,

Deskription

„Welche
Konsumertypen
gibt es?“

Vorhersage

„Sagen
Facebook-Likes
die Persönlichkeit
vorher?“

Kausalität

„Erhöht
Achtsamkeits-
training die
Konzentration?“

Wissenschaft ist meist an kausalen Fragen interessiert

Deskription

Vorhersage

Kausalität

Kausalität ist cool.

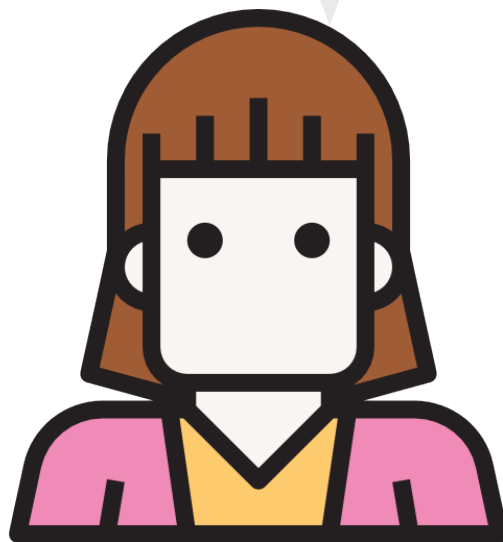


Wolfi

Journal of Applied Psychology mag auch Kausalforschung

Ausgabe 4 in 2020

10 von 12 Studien
der Ausgabe
verwenden Kausal-
Sprache.



Angi

Title	quotes (abstract)	causal language?
The generation and function of moral emotions in teams: An integrative review.	„influence on individual team members' moral emotions“	yes
On melting pots and salad bowls: A meta-analysis of the effects of identity-blind and identity-conscious diversity ideologies.	„improve intergroup relations“ „the effects of identity-blind ideologies“	yes
Political affiliation and employment screening decisions: The role of similarity and identification processes.	„to examine the effects of“	yes
A dynamic account of self-efficacy in entrepreneurship.	„self-efficacy energizes action because“	yes
Coworker support and its relationship to allostasis during a workday: A diary study on trajectories of heart rate variability during work.	„We examined the effect of“	yes
A theoretical assessment of dismissal rates and unit performance, with empirical evidence.	"utility analysis suggests that increasing dismissal rates can improve performance“	yes
Motivation to lead: A meta-analysis and distal-proximal model of motivation and leadership.	„the three MTL types partially explained the relationship“	no
Putting leaders in a bad mood: The affective costs of helping followers with personal problems.	„how such helping acts may impact leaders“ „leaders with high (vs. low) managerial experience were less affected by“	yes
When goals are known: The effects of audience relative status on goal commitment and performance.	„investigating how the perceived relative status of a goal audience influences goal commitment“	yes
Selecting response anchors with equal intervals for summated rating scales.		no
It hurts me too! (or not?): Exploring the negative implications for abusive bosses.	„we propose that perpetrated abuse impacts these supervisor outcomes“	yes
How can employers benefit most from developmental job experiences? The needs-supplies fit perspective.	„developmental job experiences (DJE) lead to positive work-related outcomes“	yes

Take-home Message der Take-home Messages



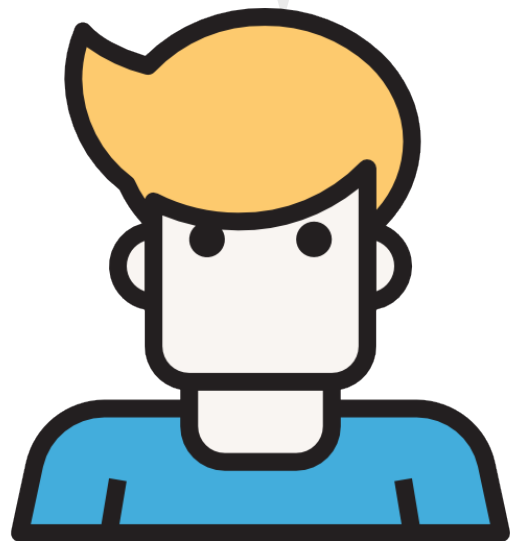
Wolfi

Glaub keiner Beobachtungsstudie.

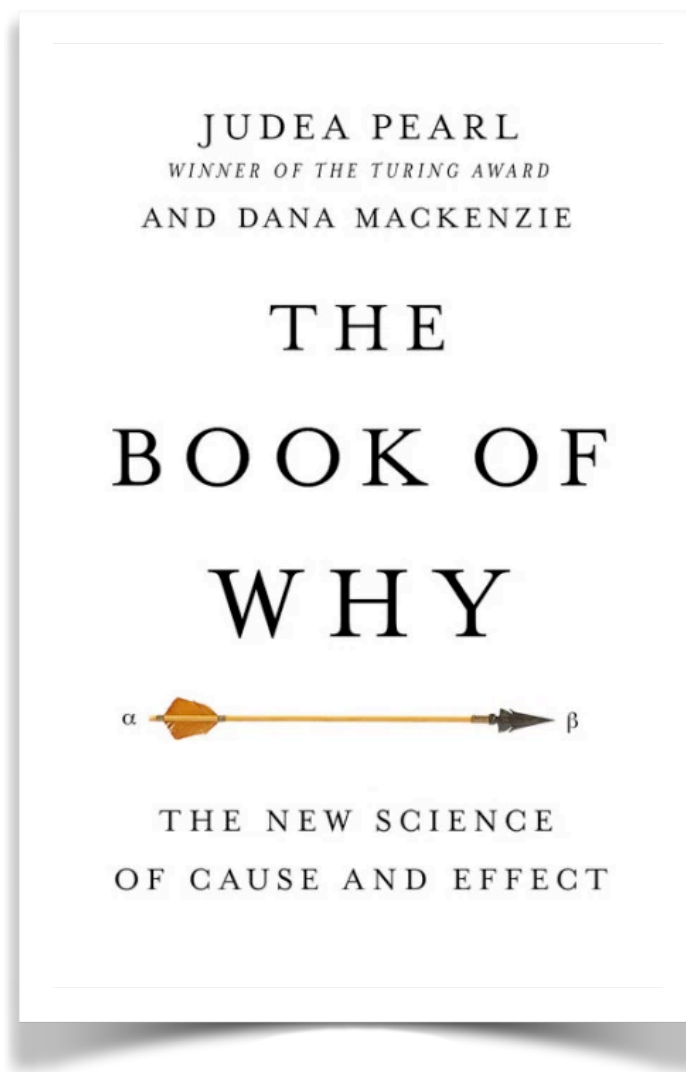
Es sei denn, sie präsentiert ein gutes
Kausalmodell.

Kausalinferenz ist ein Juwel für die Statistik

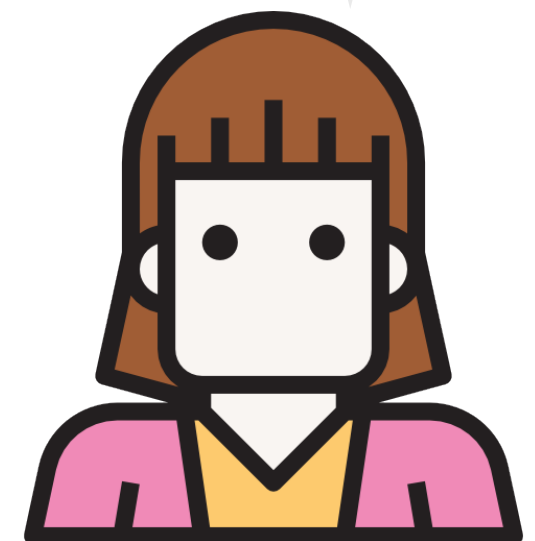
Das erzählt doch bloß ein verrückter Prof!



Don



Hm, aber Judea Pearl hat dafür den Turing Award erhalten.



Angi

Einstiegsliteratur



- Elwert, F. (2013). Graphical causal models. In S. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 245–273). Springer. https://www.researchgate.net/publication/278717528_Graphical_Causal_Models
- Lübke, K., Gehrke, M., Horst, J., & Szepannek, G. (2020). Why We Should Teach Causal Inference: Examples in Linear Regression with Simulated Data. *Journal of Statistics Education*, 1–17. <https://doi.org/10.1080/10691898.2020.1752859>
- Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>

- Corvetti, C. (2006). Saratoga Houses. <https://rdr.io/cran/mosaicData/>
- Dablander, F. (2020). An Introduction to Causal Inference [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/b3fkx>
- Dedering, U. (2010). Map of the USA [Map]. https://en.wikipedia.org/wiki/Saratoga_Springs,_New_York#/media/File:Usa_edcp_relief_location_map.png
- Elwert, F. (2013). Graphical causal models. In S. Morgan (Ed.), Handbook of causal analysis for social research (pp. 245–273). Springer. https://www.researchgate.net/publication/278717528_Graphical_Causal_Models
- Hernán, M. A., Hsu, J., & Healy, B. (2019). A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *Chance*, 32(1), 42–49. <https://doi.org/10.1080/09332480.2019.1579578>
- item2101. (2020). Avatar Icon Pack [Icon]. www.flaticon.com. <https://www.flaticon.com/packs/avatar-14?k=1587995971688>
- Lübke, K. (2020, February). Introduction to Causal Inference. Dozententage der FOM, Essen.
- Lübke, K., Gehrke, M., Horst, J., & Szepannek, G. (2020). Why We Should Teach Causal Inference: Examples in Linear Regression with Simulated Data. *Journal of Statistics Education*, 1–17. <https://doi.org/10.1080/10691898.2020.1752859>
- Pearl, J. (2009). *Causality*. Cambridge university press.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect* (First edition). Basic Books.
- Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289–310. <https://doi.org/10.1214/10-STS330>
- UpstateNYer. (2009). Saratoga County, New York, USA,. https://en.wikipedia.org/wiki/Saratoga_Springs,_New_York#/media/File:Downtown_Saratoga_Springs.jpg