

Een onnodig gecompliceerde titel

Misja Mikkers & Gertjan Verhoeven

Nederlandse Zorgautoriteit (NZa) & Tilburg University

31 oktober 2018

Inleiding

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Voorbeeld van een dataset

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

X_1	X_2	...	X_i	I	Y0	Y1
Man	9	14	1	0	67	NA
Vrouw	60	36	0	1	NA	113
Vrouw	7	2	1	1	NA	54
...

Voorspellen versus begrijpen

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

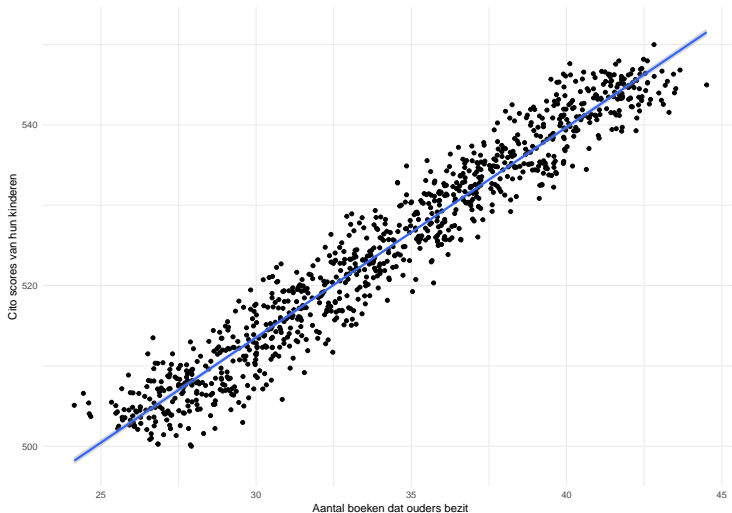
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Voorspellen versus begrijpen

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

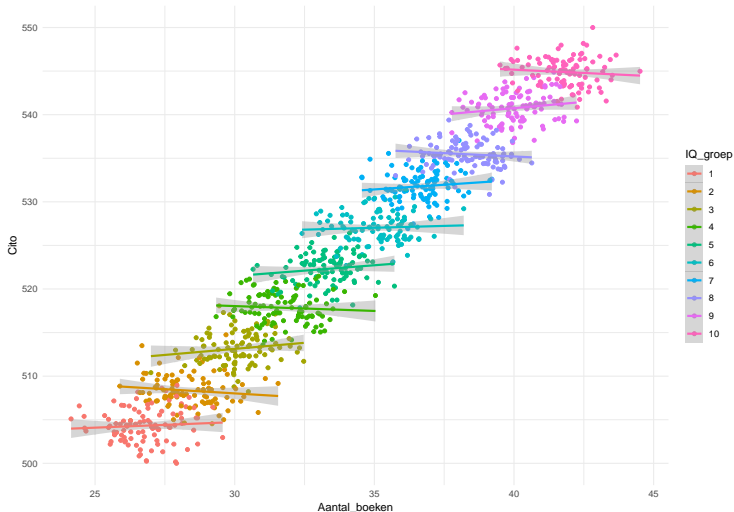
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Basisvormen DAGs

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

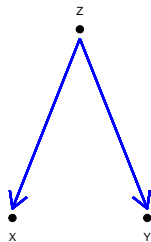
Causaliteit

DAGS

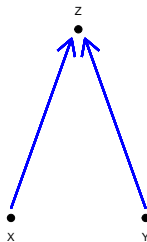
Machine learning
en causaliteit

Conclusie

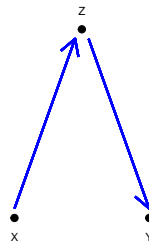
Confounder



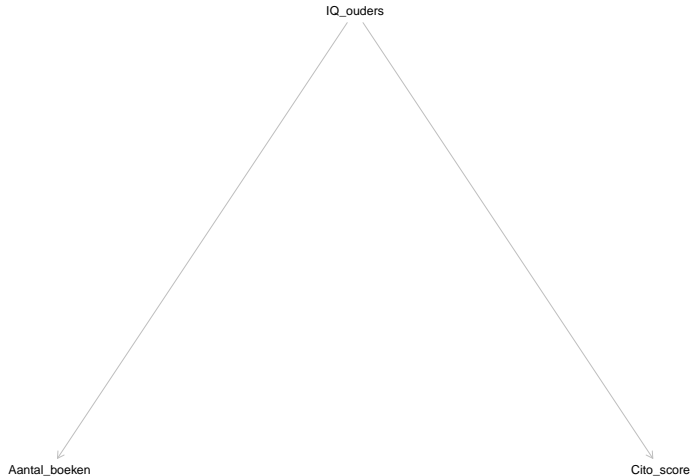
Collider



Mediator



Confounder



Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

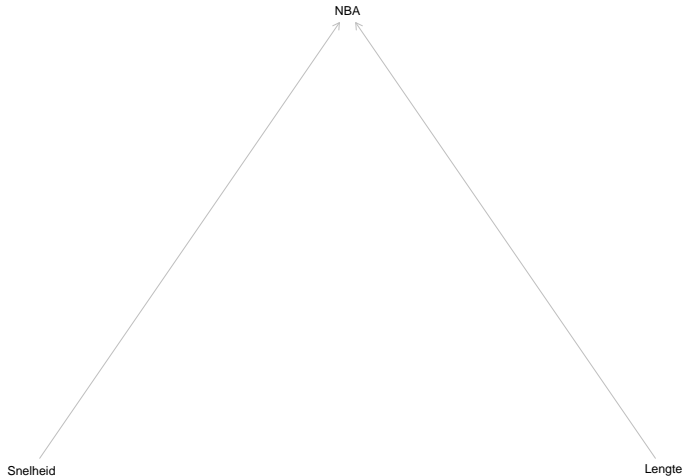
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Veroorzaakt Lengte Snelheid?

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

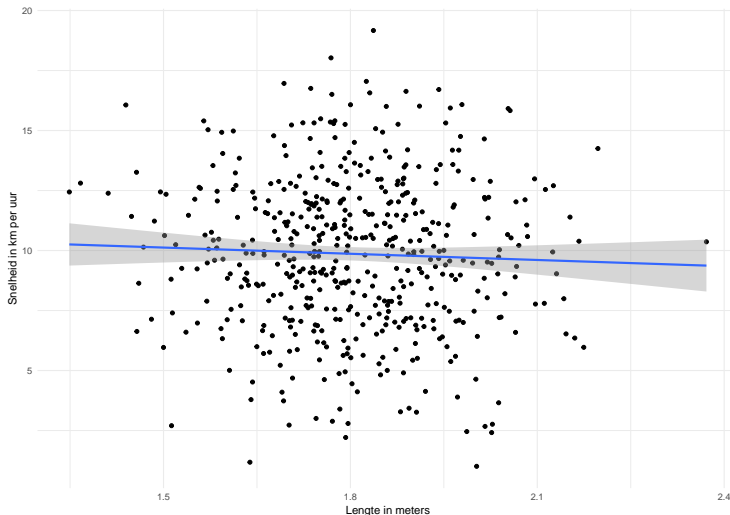
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Veroorzaakt Lengte Snelheid?

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

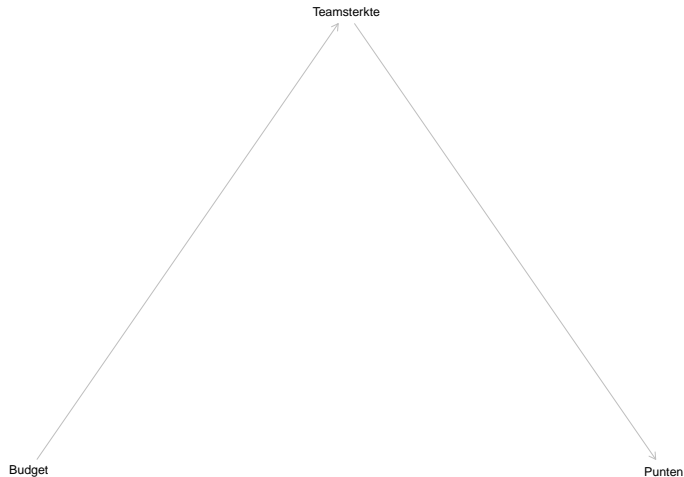
DAGS

Machine learning
en causaliteit

Conclusie



Mediator



Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

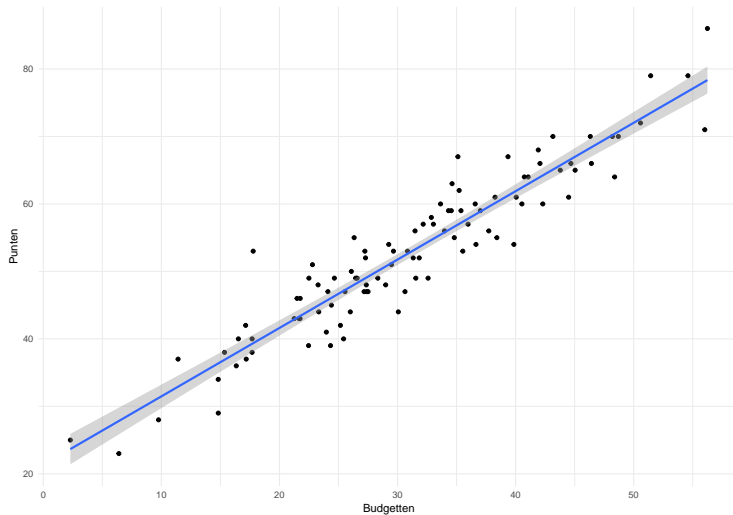
Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Budgetten en punten



Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Budgetten en punten

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

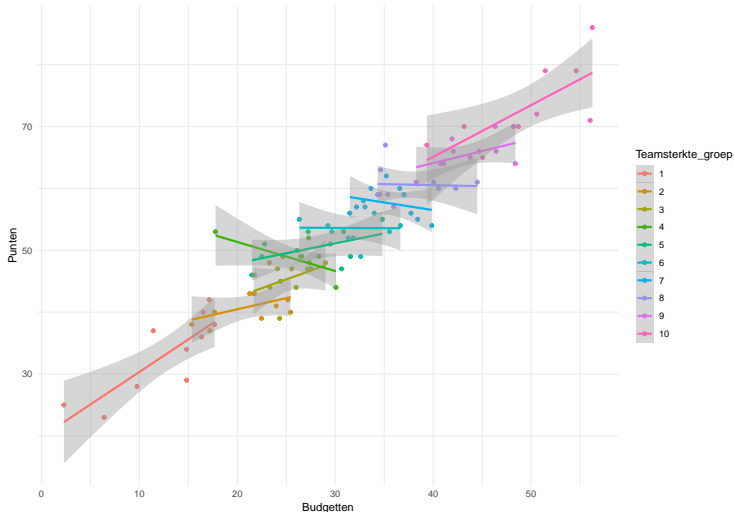
Inleiding

Causaliteit

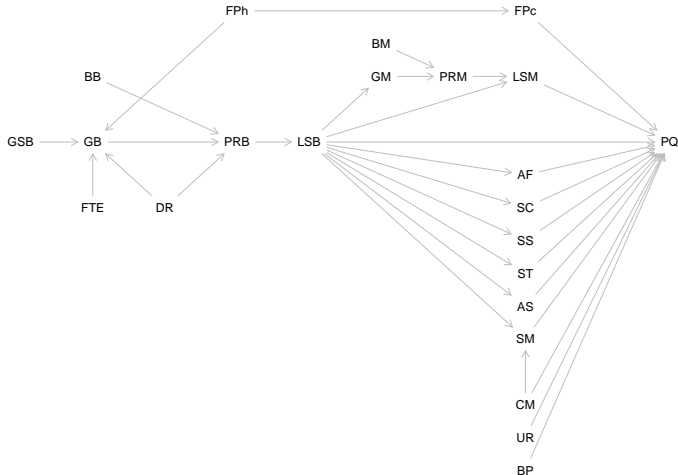
DAGS

Machine learning
en causaliteit

Conclusie



Voorbeeld van een meer complexe DAG



Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie

Machine learning en causaliteit

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

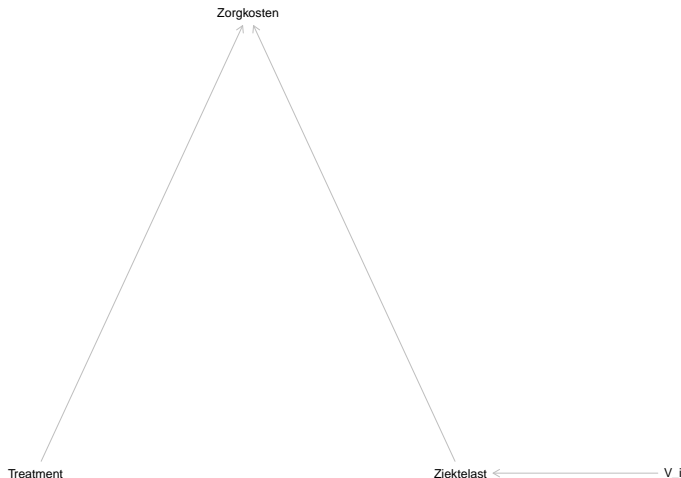
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



De variabele *Ziektelast* is bepaald op basis van de volgende formule:

$$\text{Ziektelast} = (1*V1+2*V2+3*V3+4*V4+5*V5+6*V6)/10000+\epsilon$$

1. Fit een Random Forest model op de data
2. Bepaal het gemiddelde treatment effect met behulp van generalized random forests (grf)

We doorlopen deze procedure 2 keer:

1. Een analyse waarbij we de ziektelast voorspellen met alle variabelen (“het verkeerde model”)
2. Een analyse waarbij we de ziektelast voorspellen met alle variabelen minus de collider *Zorgkosten* (“het goede model”)

Summary statistics

Statistic

N

Mean

St. Dev.

Min

Pctl(25)

Pctl(75)

Max

Ziektelast

1,000

1.050

0.286

Voorspellingen Random Forest

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

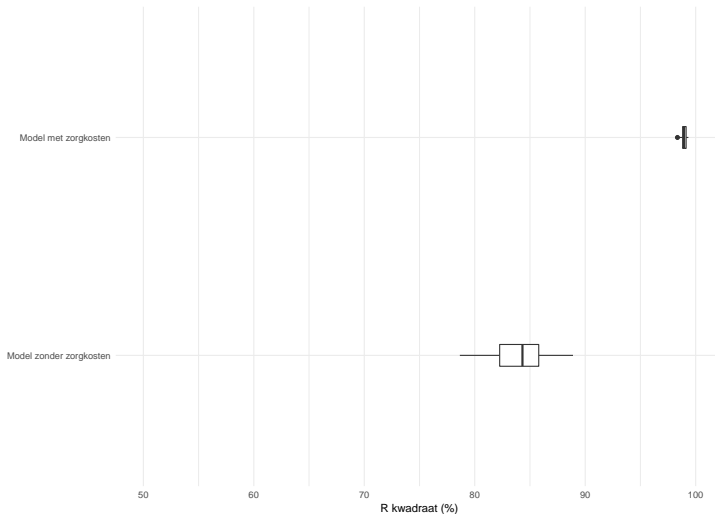
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



Schatting van het treatment effect

Een onnodig
gecompliceerde
titel

Misja Mikkers &
Gertjan
Verhoeven

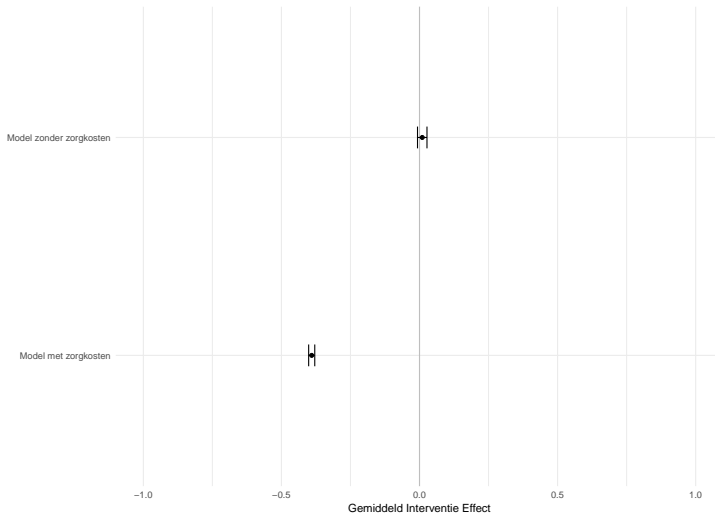
Inleiding

Causaliteit

DAGS

Machine learning
en causaliteit

Conclusie



- ▶ Het is essentieel om een causaal model te hebben
- ▶ Wanneer machine learning wordt gebruikt is de verleiding groot om alle variabelen in het model te gebruiken
- ▶ We hebben laten zien dat dit kan leiden tot misleidende conclusies