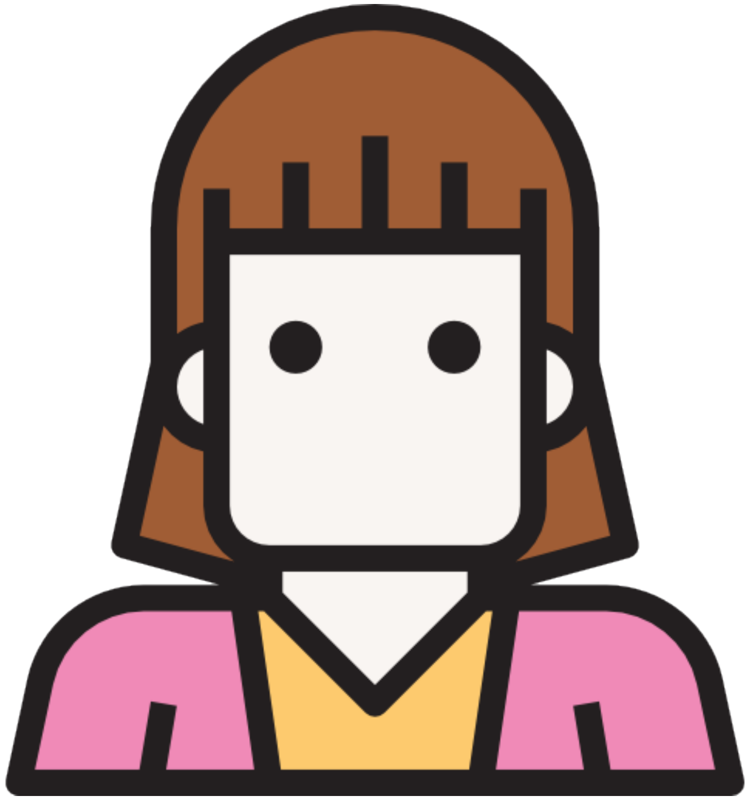
Why most published research findings are of little use for business decisions

An introduction to causal modelling

Sebastian Sauer

Case Study: Angela's got a new job as market researcher

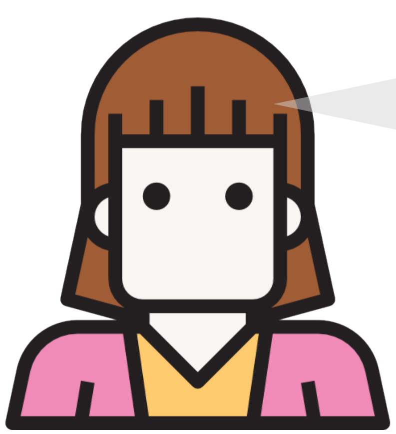


Angela M.,
Market Researcher

Image source: Flaticon

Saratoga County, one of the top places to live at in the USA

Yeah, I love my new job!

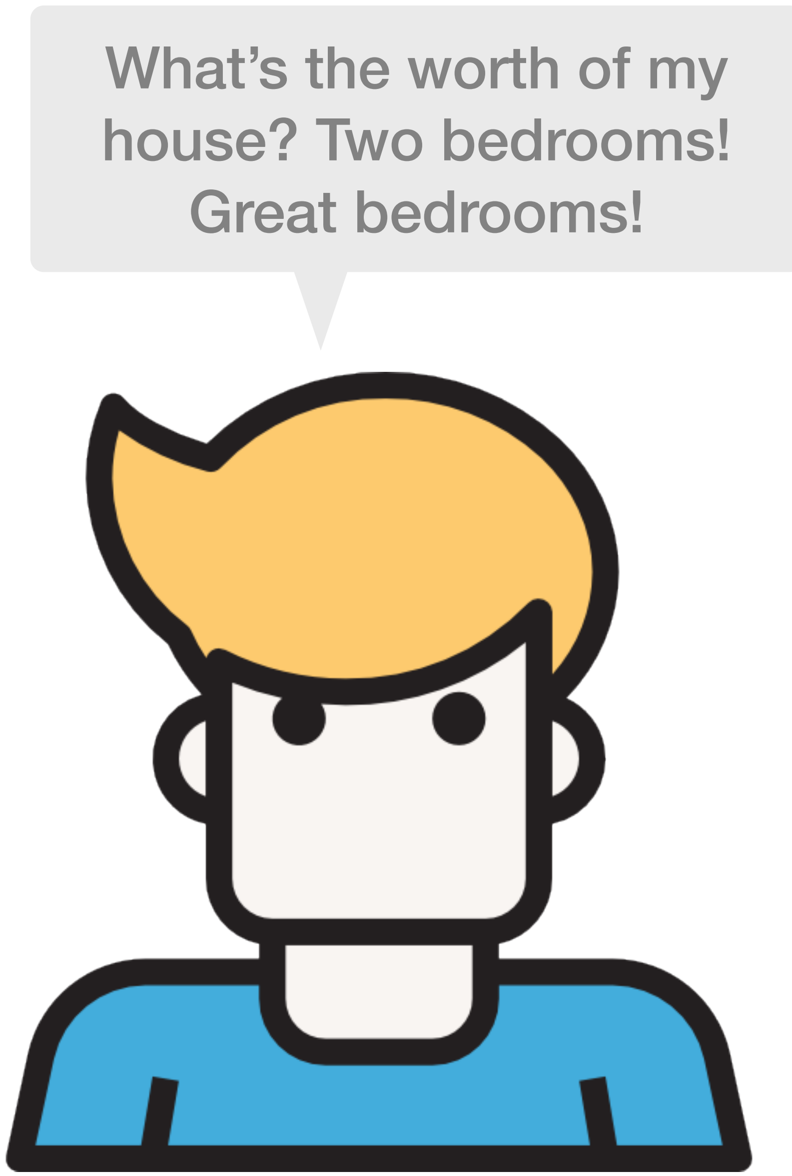


Angi

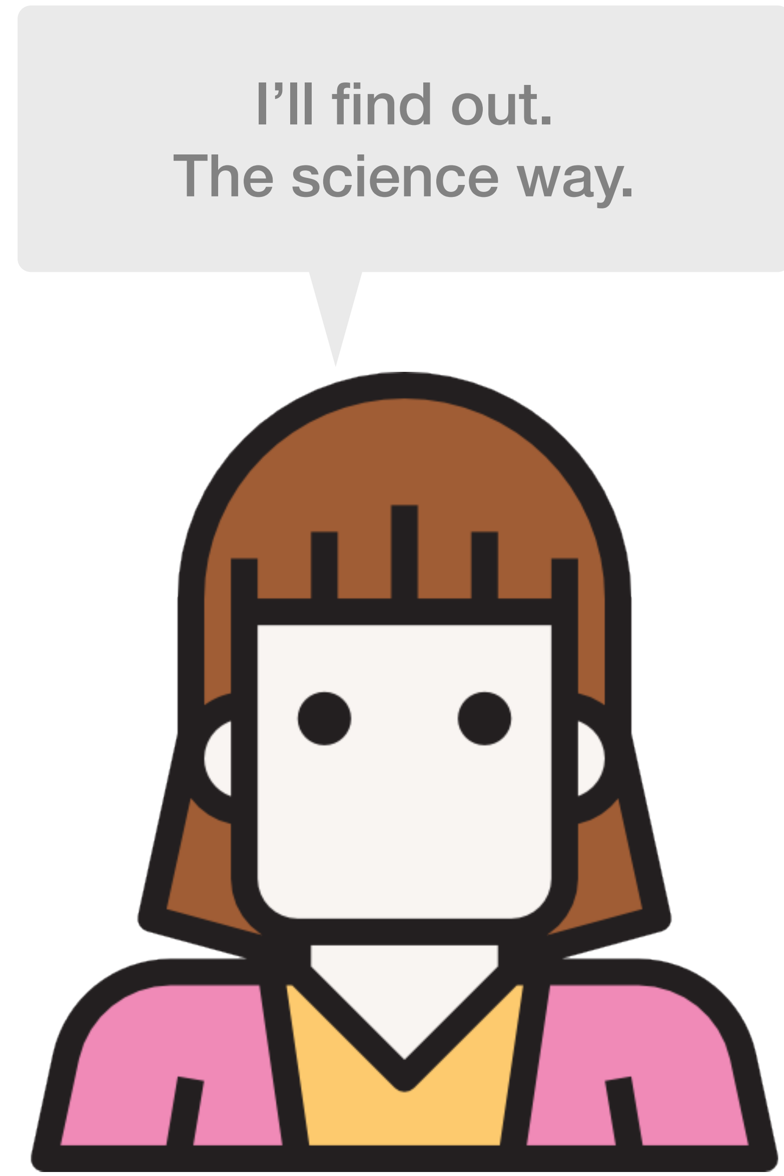
Image source: Dedering, UpstateNYer

Angi’s first project in the new job

Predicting the price of a two-bedroom house



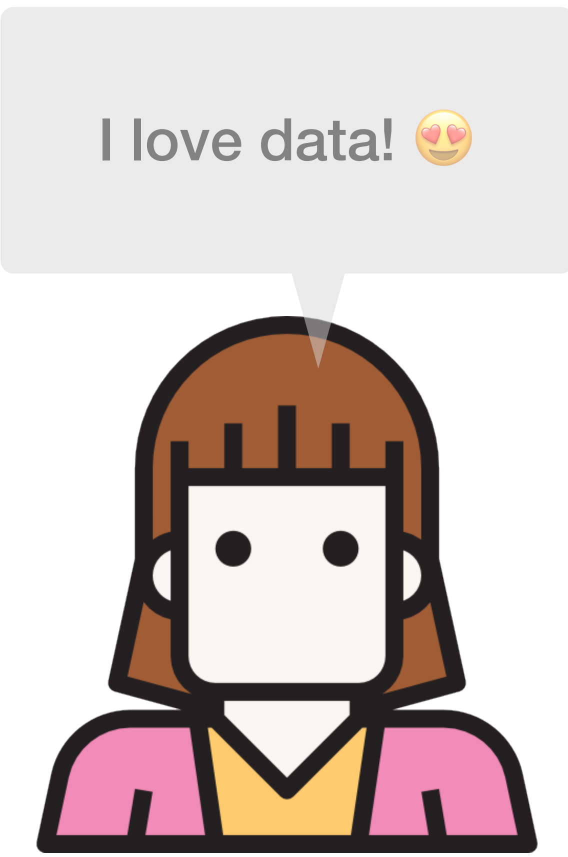
Don, real estate giant



Angi, market researcher

Here’s a glimpse on her data

First 10 rows of 1728



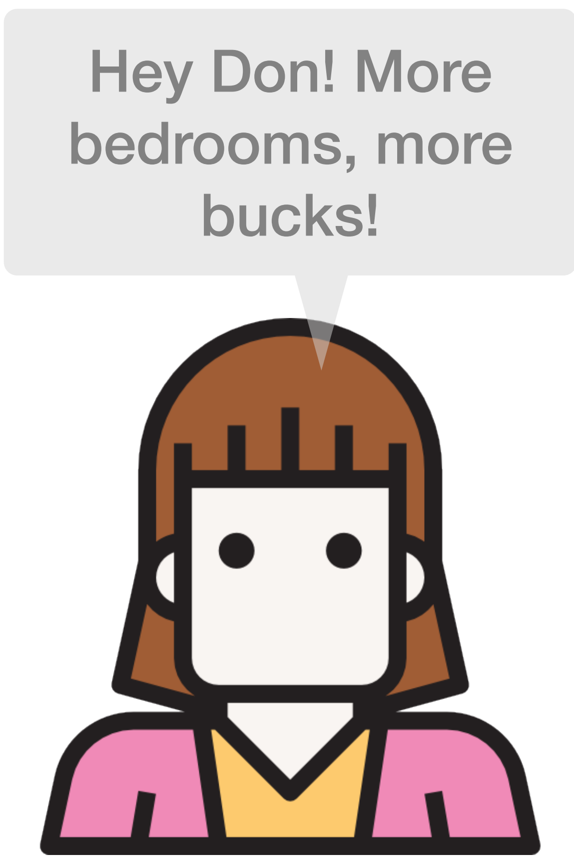
Angi

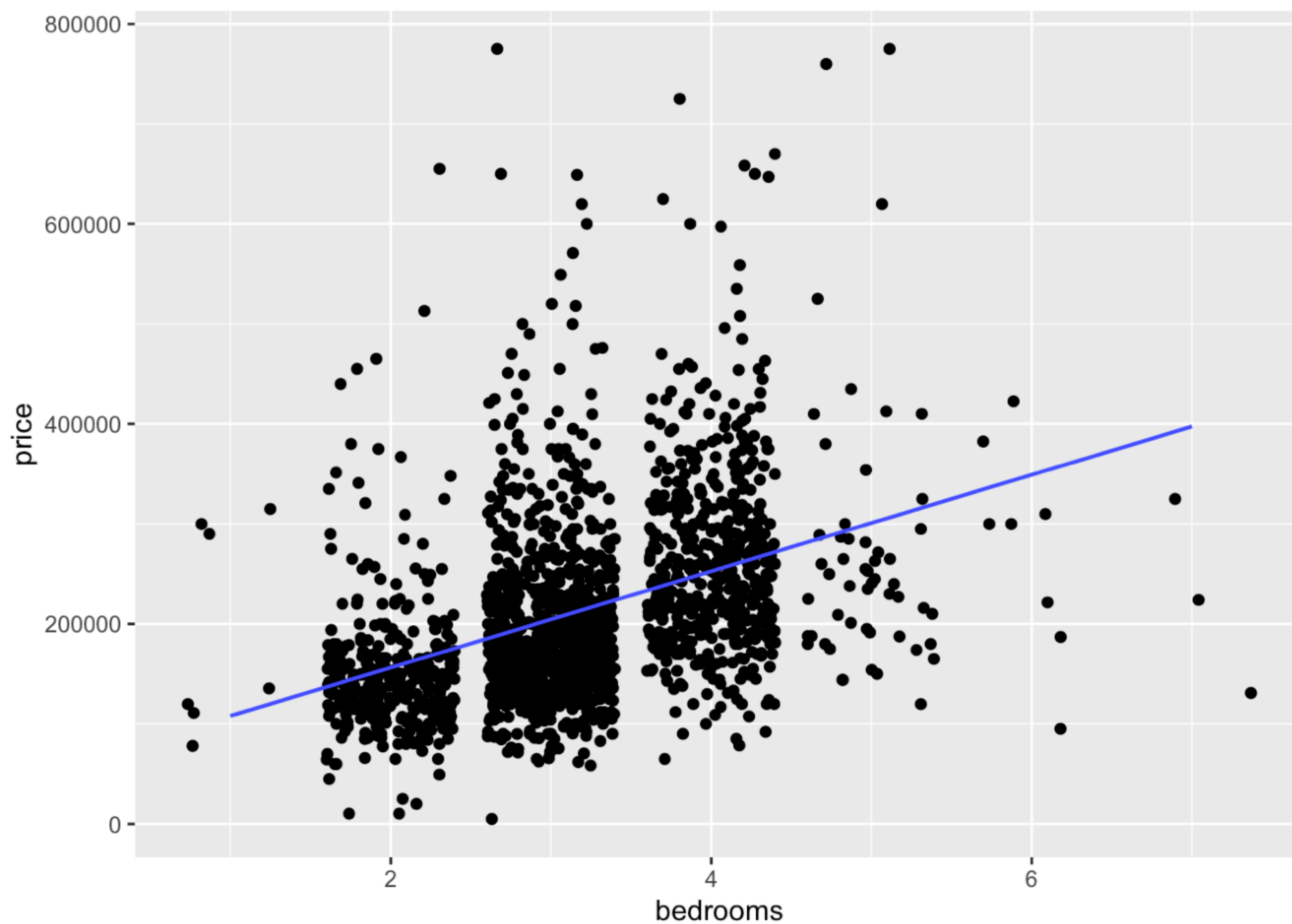
price	square-feet	age	a/c	fireplace	heating	...
132500	84.17	42	No	Yes	Electricity	
181115	181.44	0	No	No	Gas	
109000	180.60	133	No	Yes	Gas	
155000	180.60	13	No	Yes	Gas	
86060	78.04	0	Yes	No	Gas	
120000	107.02	31	No	Yes	Gas	
153000	255.67	33	No	Yes	Oil	
170000	154.40	23	No	Yes	Oil	
90000	151.62	36	No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

Corvetti, C. (2006). *Saratoga Houses*. https://rdrr.io/cran/mosaicData/

Model 1: Price as a function of number of bedrooms

The more bedrooms, the higher the price of the real estate





Corvetti, C. (2006). *Saratoga Houses*. https://rdrr.io/cran/mosaicData/

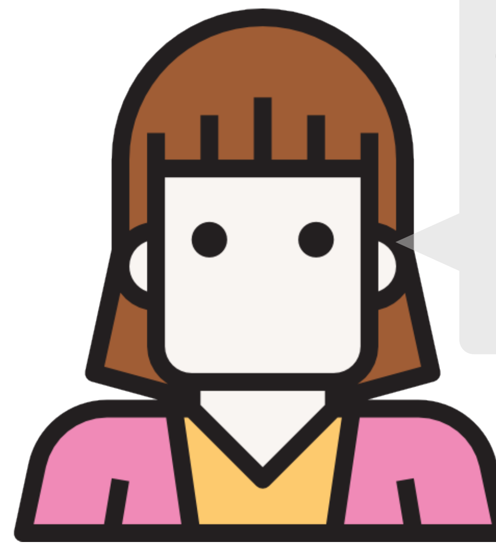
Angi reporting back to Don. Don isn't happy.

```
model1 <- lm(price ~ bedrooms,
data = SaratogaHouses)
coef(model1)
```

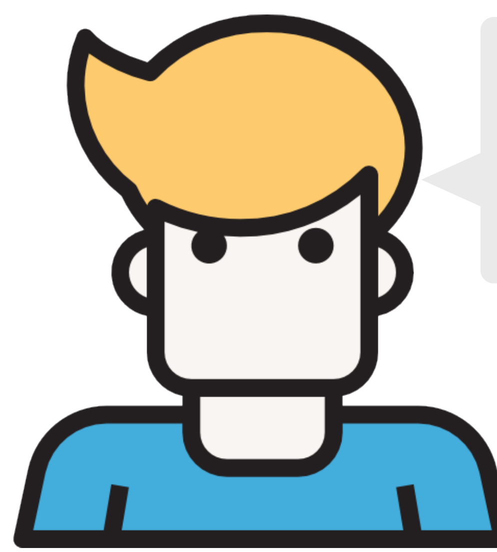
```
## (Intercept)    bedrooms
##    59862.96    48217.81
```

```
dons_house <- data.frame(bedrooms = 2)
predict(model1, dons_house)
```

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

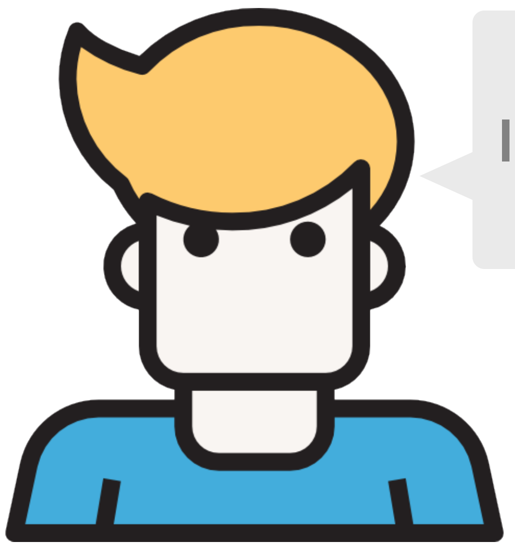


Angi



Don

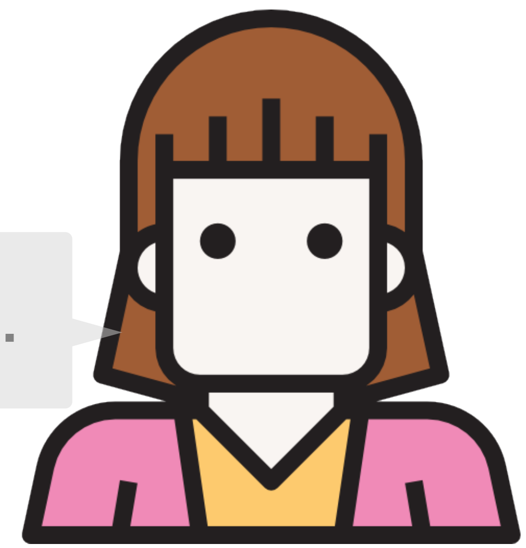
Don’s got an idea: Split each bedroom into two



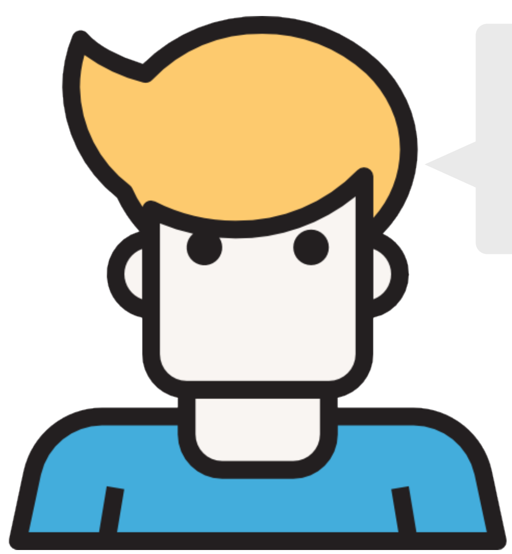
I’ll split each bedrooms into two!

Don

Wait …



Angi



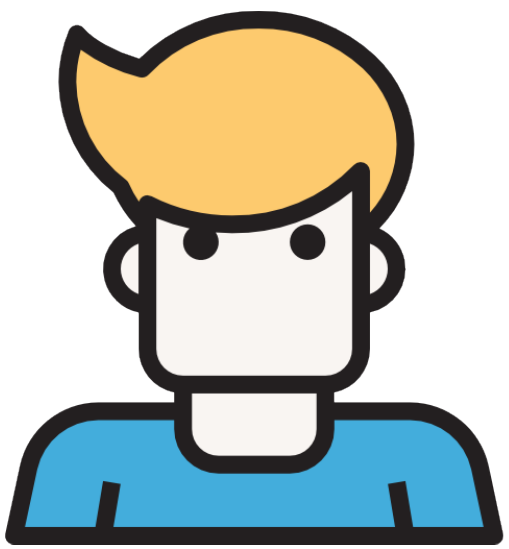
Crunch the data – now!

With 4 bedrooms, the price rises to 250k, model 1 says

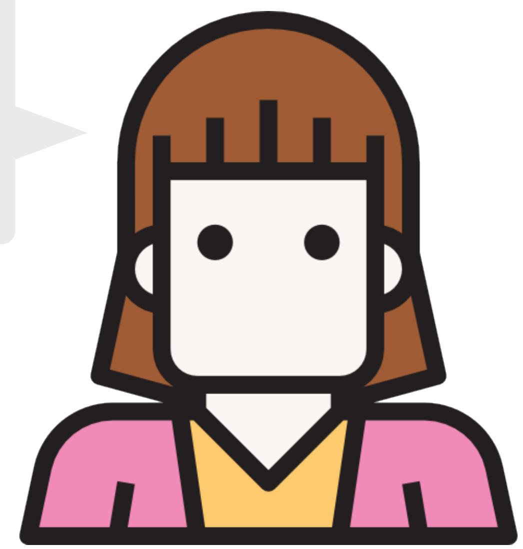
House price with four bedrooms

```
dons_new_house <- data.frame(bedrooms = 4)
predict(model1, dons_house)
```

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



Don



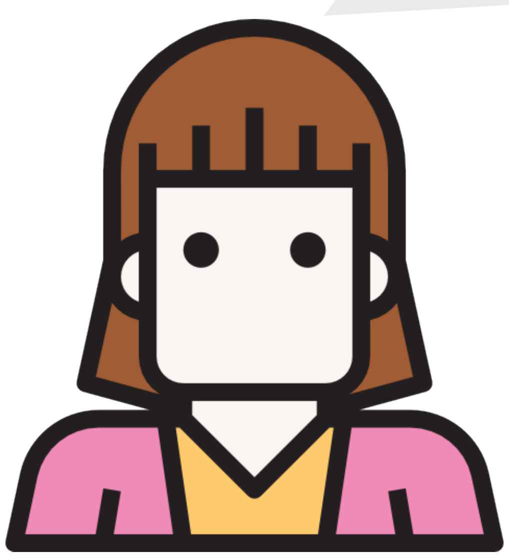
Angi

Model 2: Price as a function of two predictors: bedrooms + living area

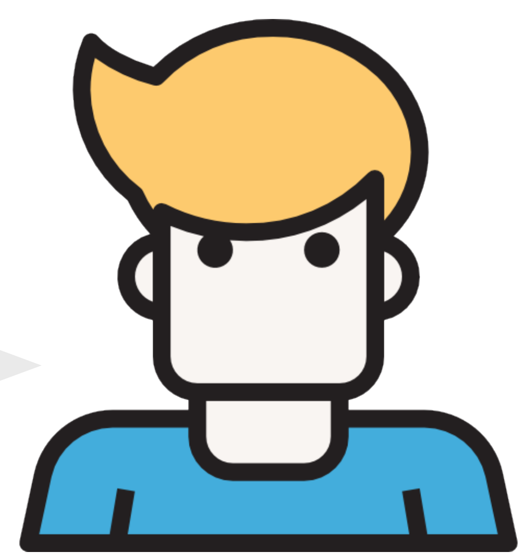
More bedrooms, *less* price – once living area is added to the model

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

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



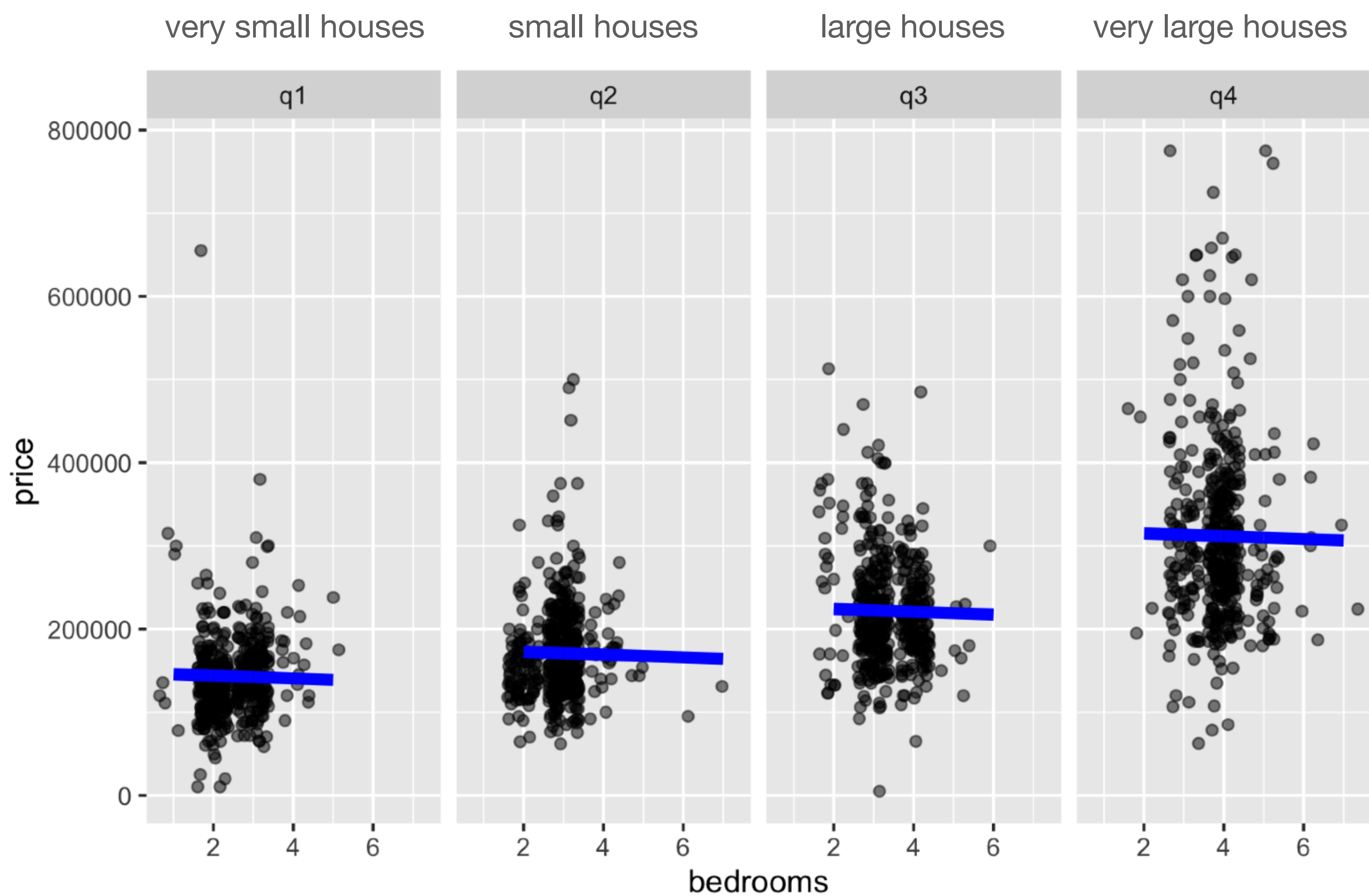
Angi



Don

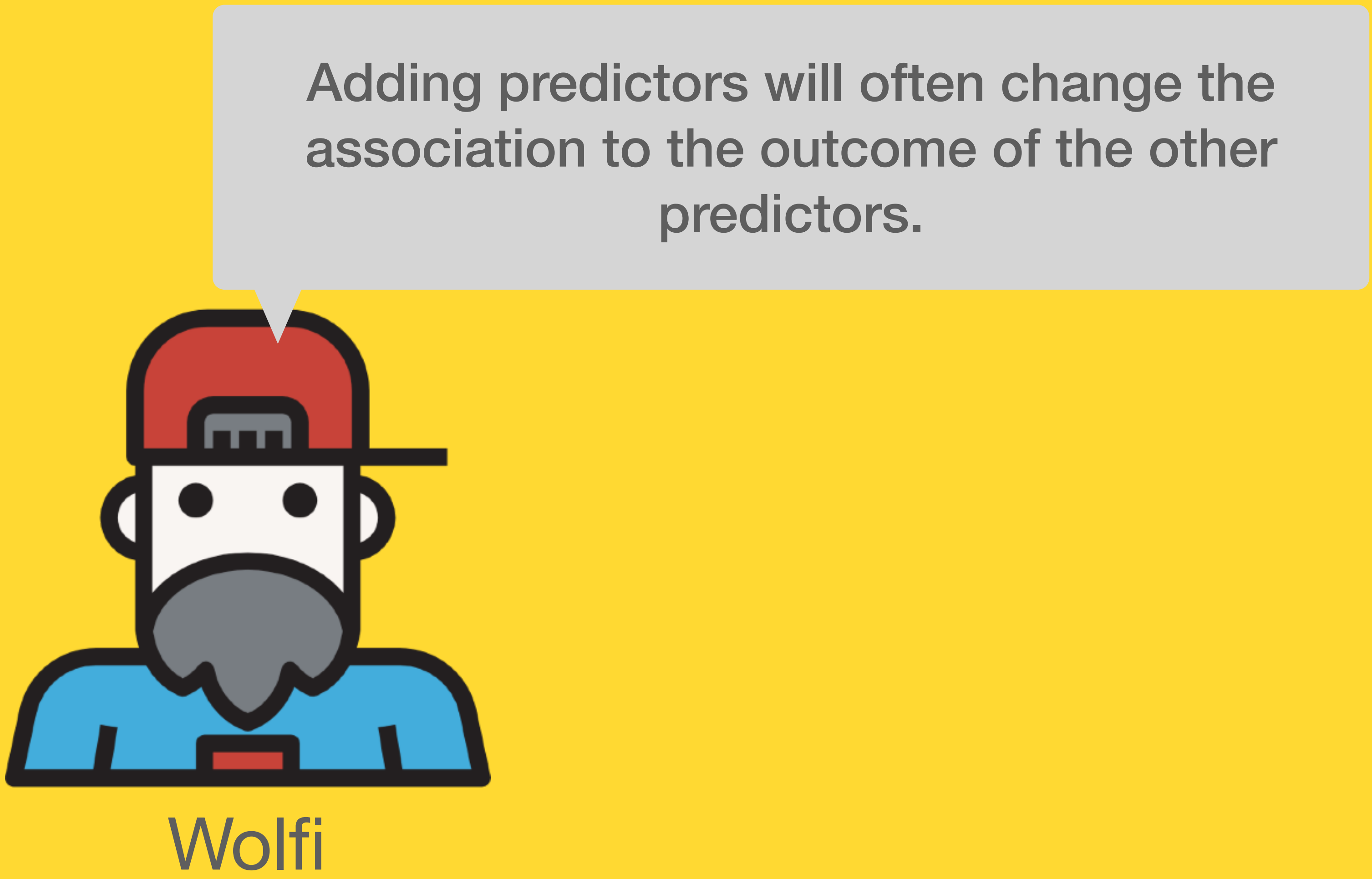
The number of bedrooms is negatively associated with price

… when the size of the living area is controlled for



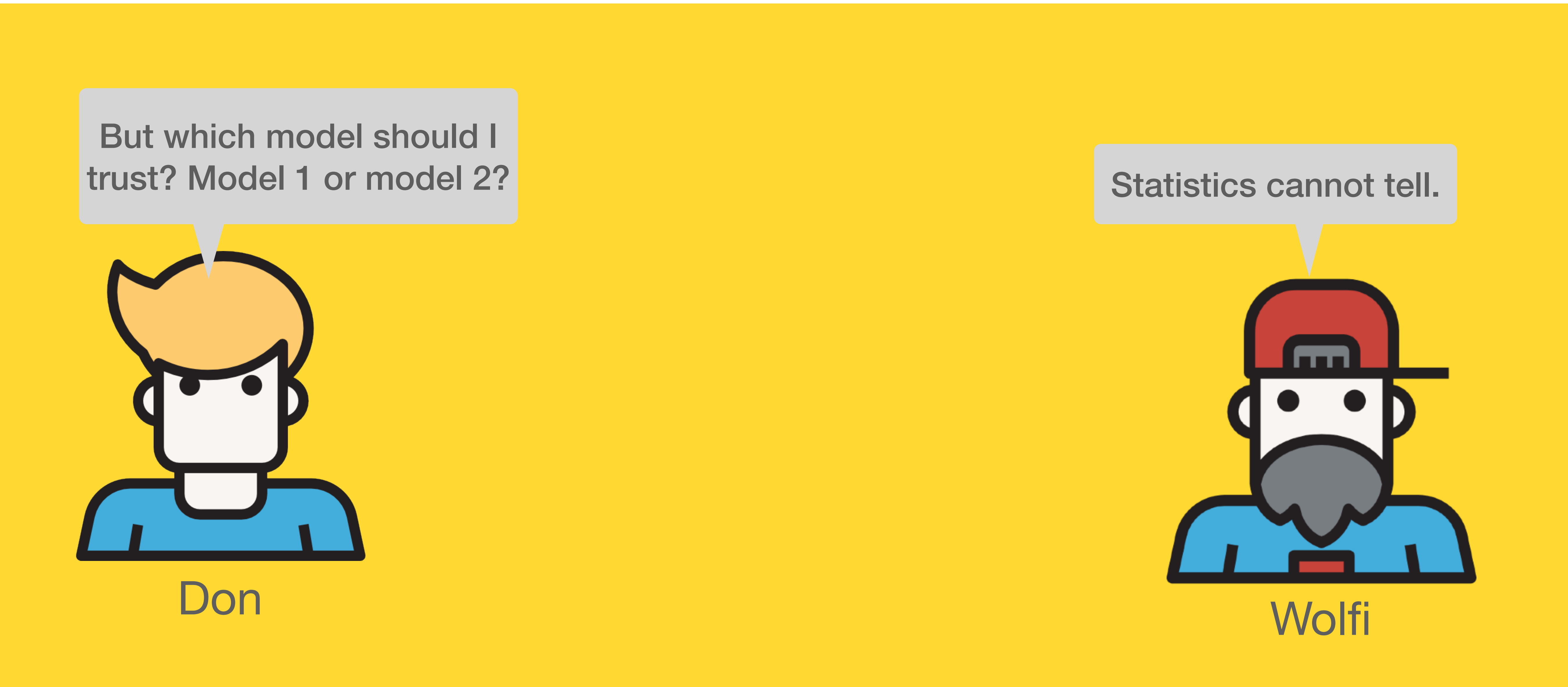
Take-home message #1

Adding predictors can (starkly) change the picture



Take-home message #2

Statistical results cannot reveal the true value

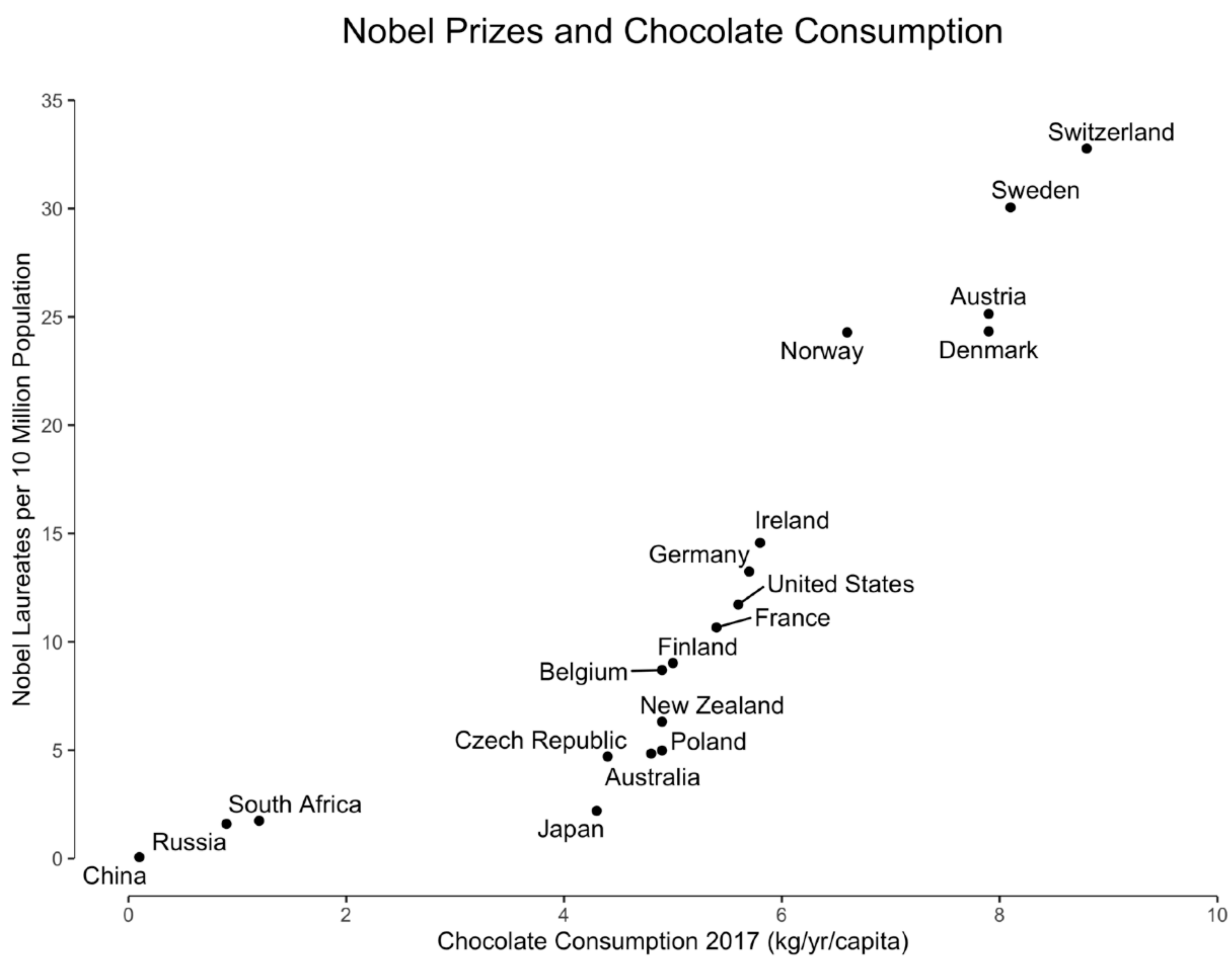


Take-home message #3

Statistical results cannot reveal the true value

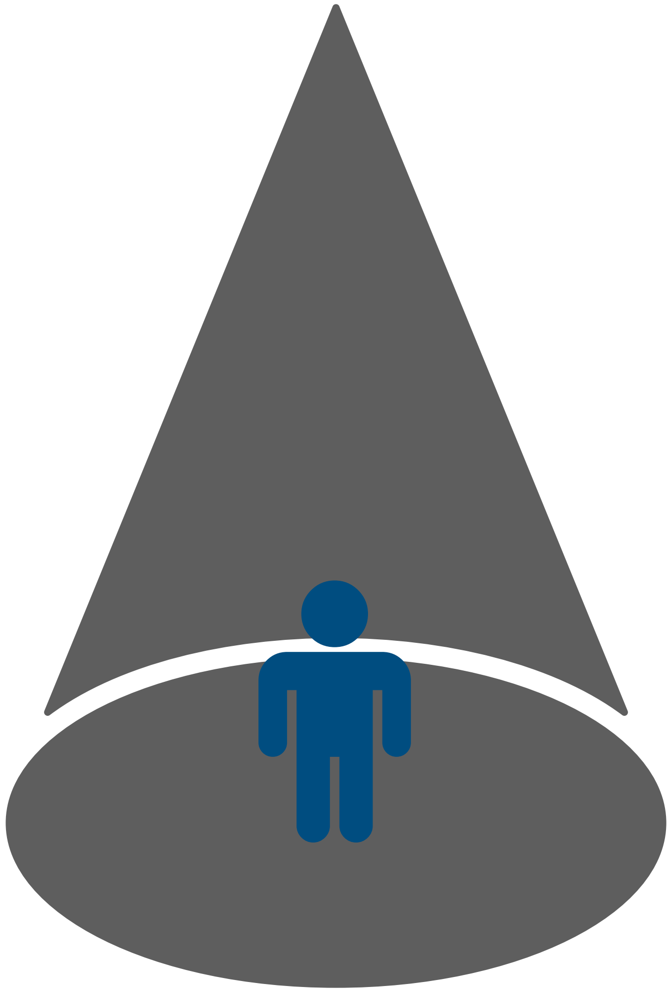
Spurious correlation: Example

Chocolate makes for Nobel prizes?!



Dablander, F. (2020). *An Introduction to Causal Inference* [Preprint]. PsyArXiv.
https://doi.org/10.31234/osf.io/b3fkw

Statistical associations can be real … or spurious

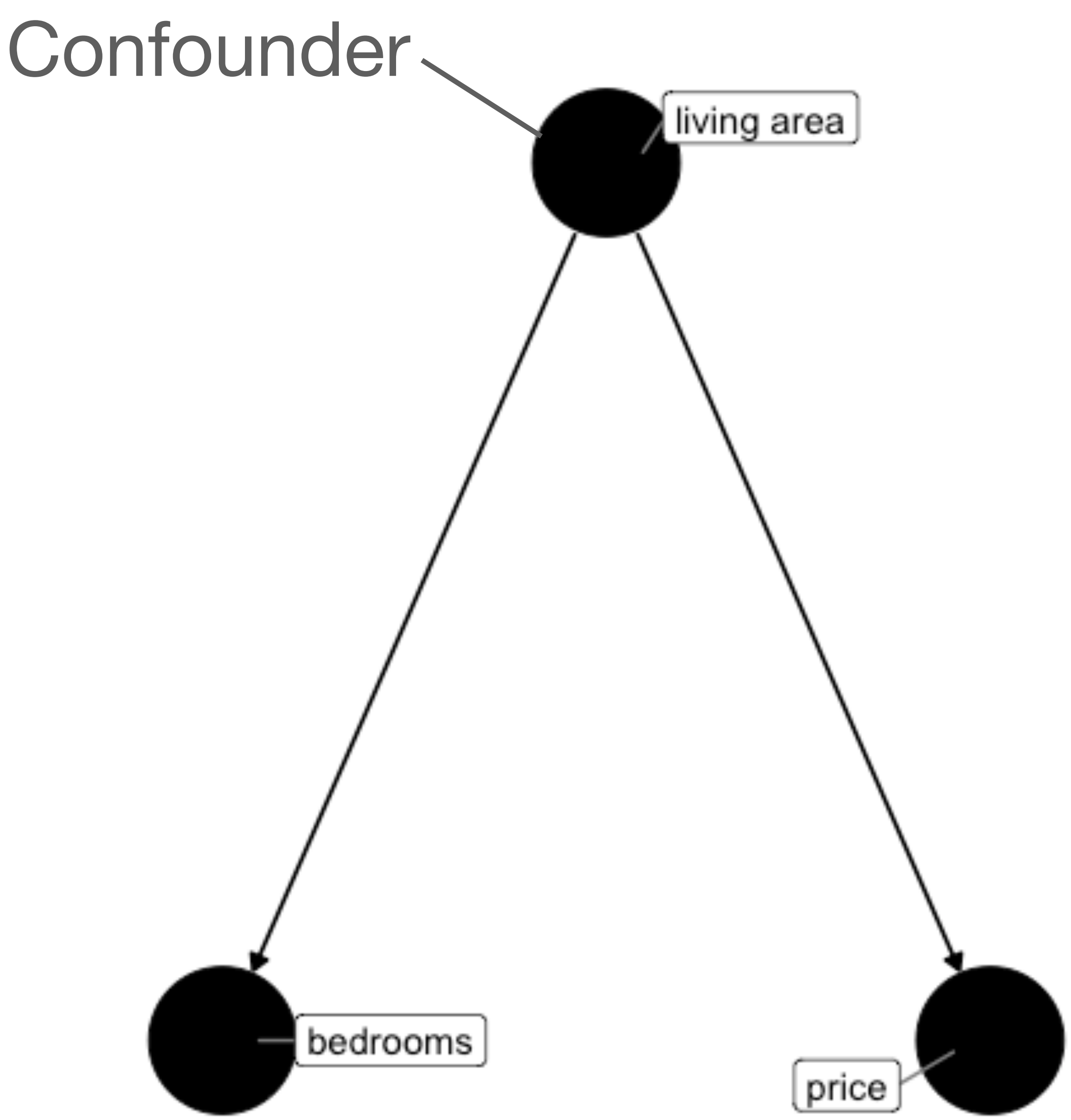


Fake (spurious)

real (causal)

Model 2: Angi's model: *livingArea* as a confounder

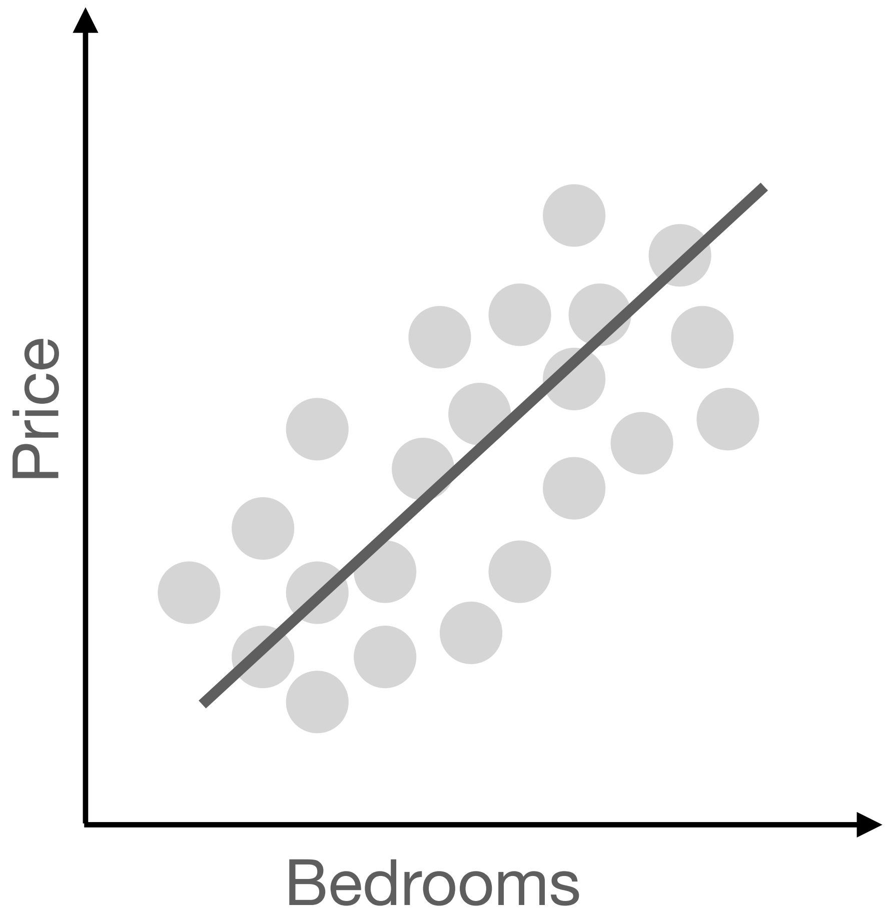
livingArea is responsible for the association of *bedrooms* and *price*



Controlling the confounder is the key

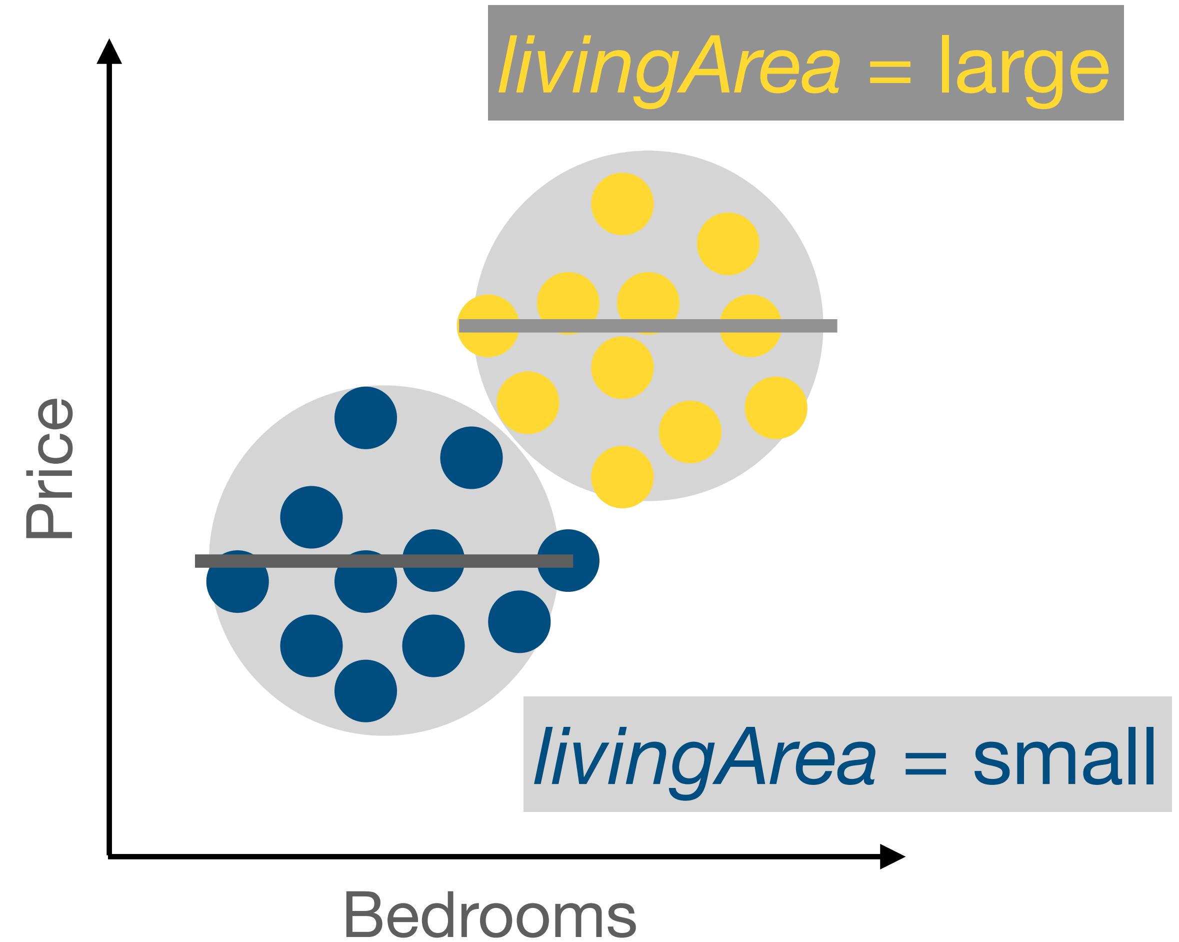
… to dissolving spurious correlation

Model 1: Confounder *livingArea* NOT controlled



Spurious correlation appears

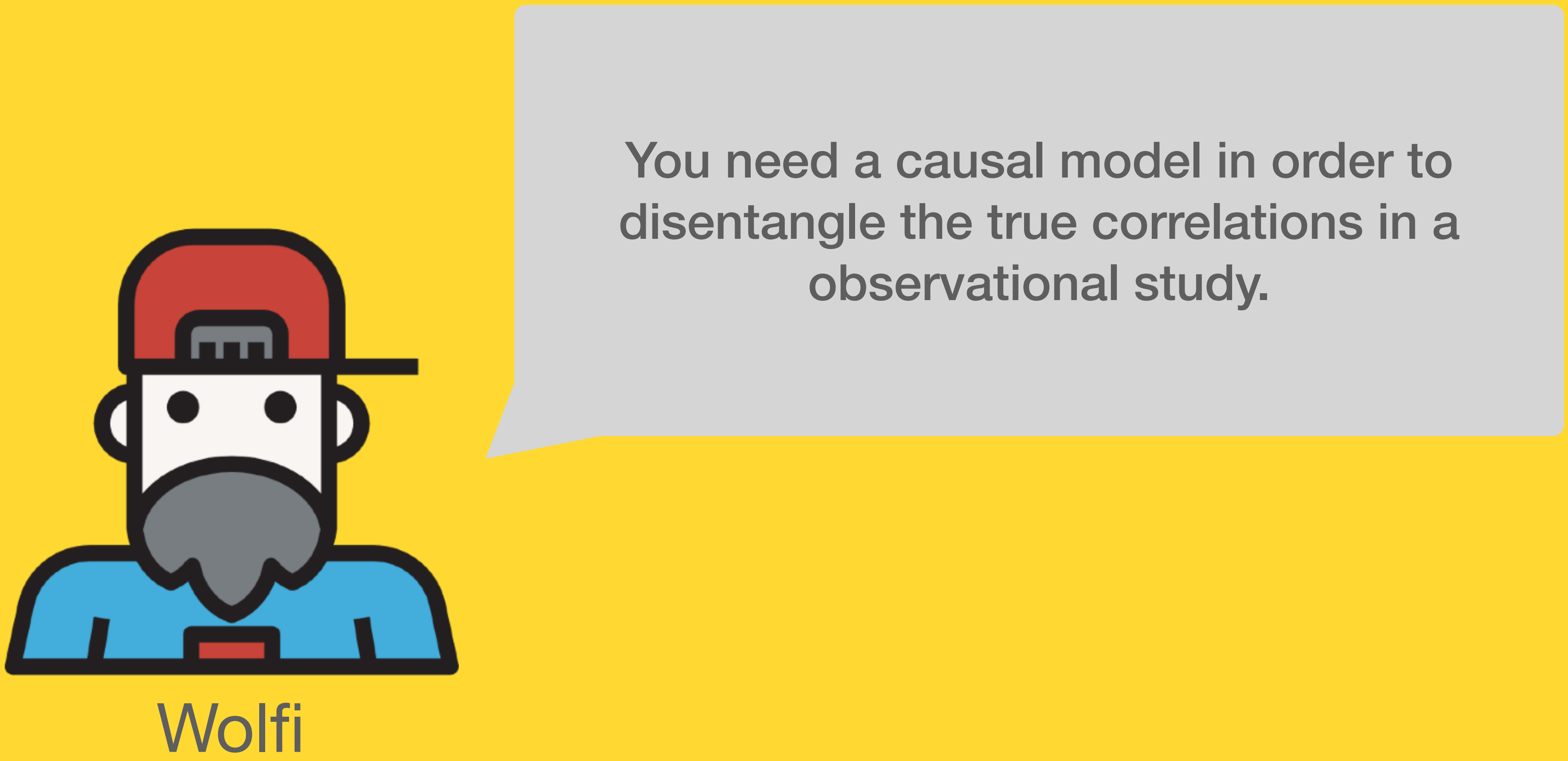
Model 2: Confounder *livingArea* controlled



Spurious correlation disappears

Take-home message #4

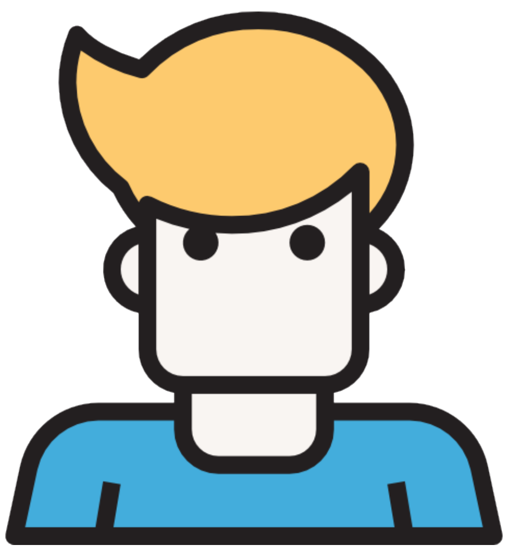
A causal model will rescue you, provided it's true



Model 1 does not fit the data

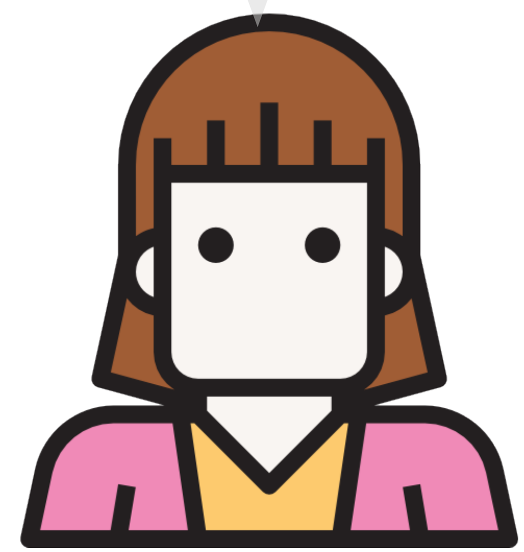
Don still likes model 1 though

Model 1 holds that *price* is influenced by *bedrooms* only. No confounders!



Don

Model 1 is falsified by the data.

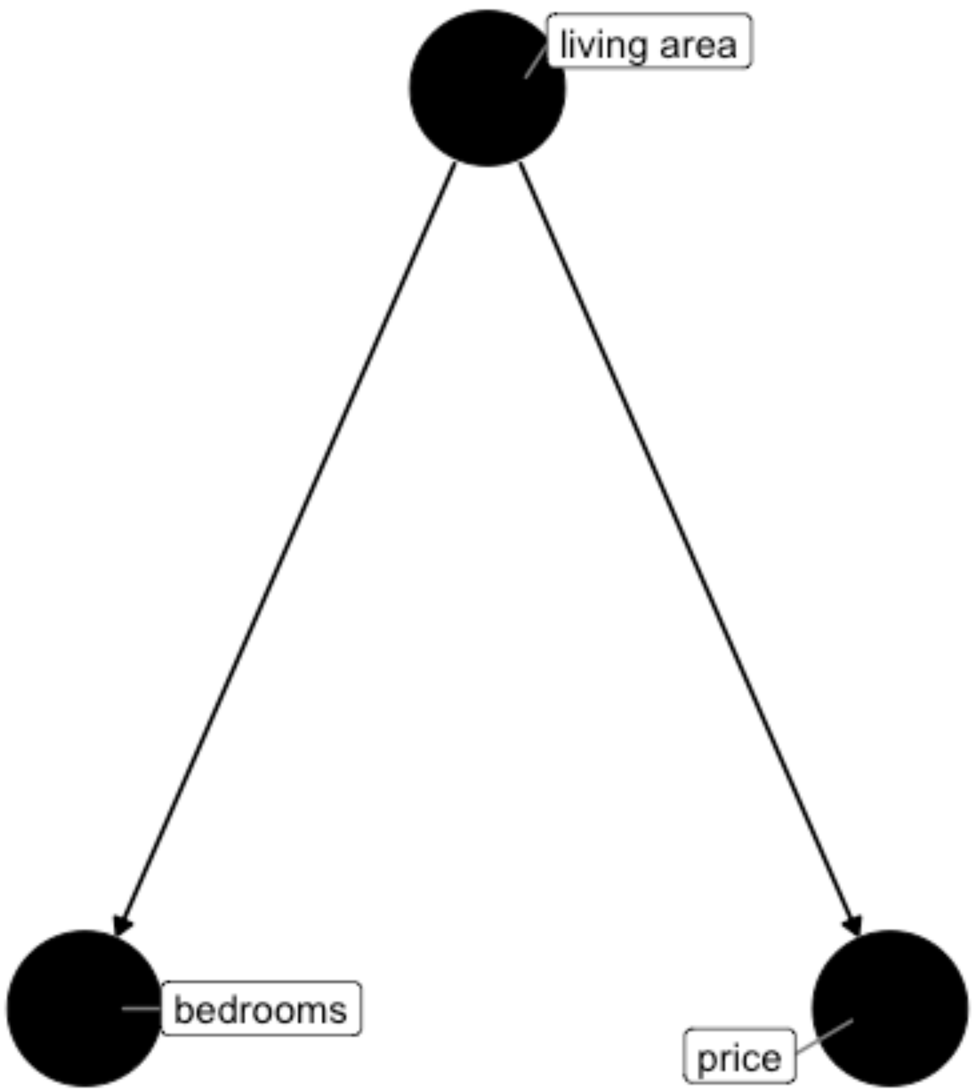


Angi

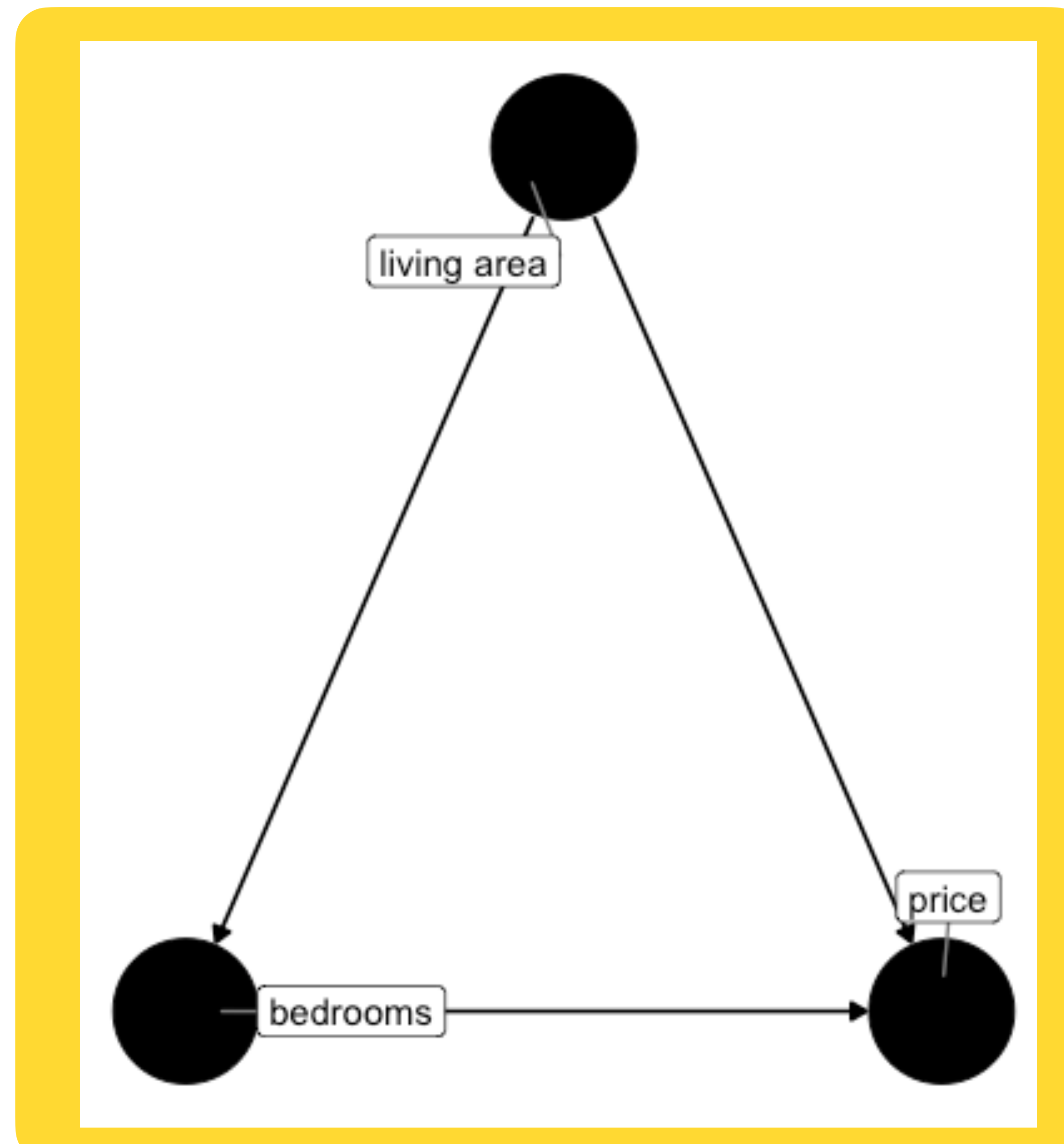
Angi’s model is not quite right either

There’s an effect of livingArea on bedrooms, as predicted by Wolfi

Angi’s model (model 2)

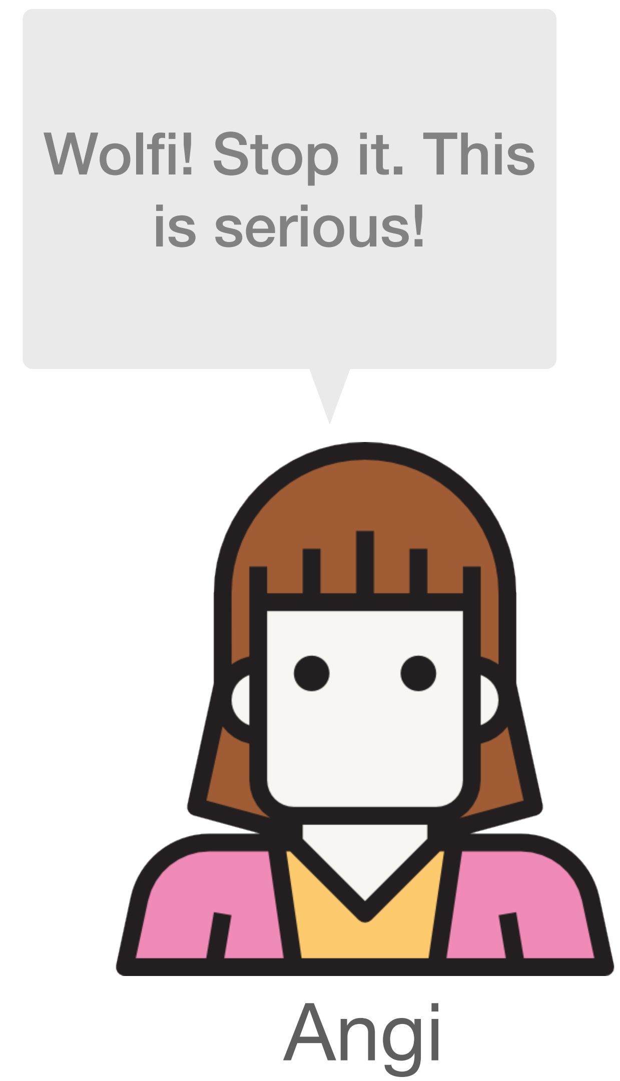


Wolfi’s model (model 3)



This model fits better than model 1 and 2.

There are three types of scientific studies



There are three types of scientific studies

descriptive

„What consumer types exist?“

associative

„Do Facebook likes predict personality?“

causal

„Does meditation increases concentration“?

„effect of X on Y“

„X impacts Y“

„X influences Y“

„X leads to Y“

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

Science is mostly concerned about causes

descriptive

associative

causal

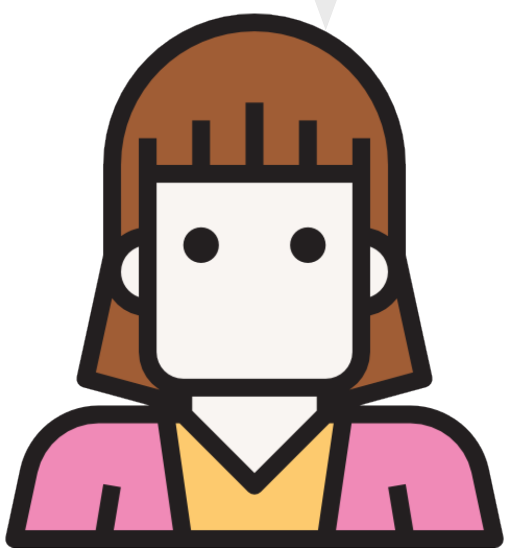
Causality is cool.

Wolfi

Journal of Applied Psychology reasons about causal questions

Indicators of causal hypotheses in recent issues (4/5 of 2020)

10 out of 12 studies presented their research questions using causal language.



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 of take-home messages

Be sceptical about advice on what to do

Don't take advice on what to do from an observational study.

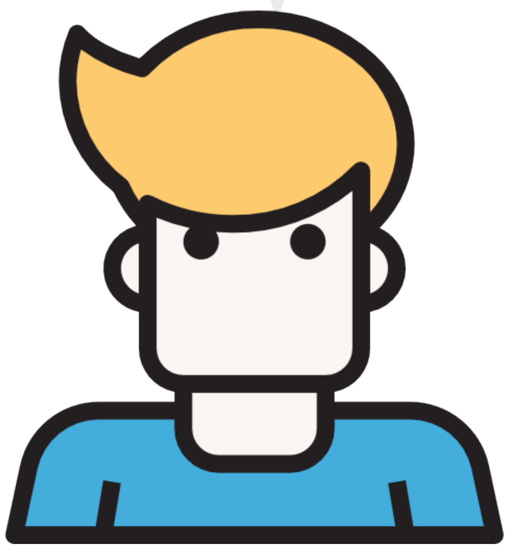
Unless is presents a convincing causal model.

Wolfi

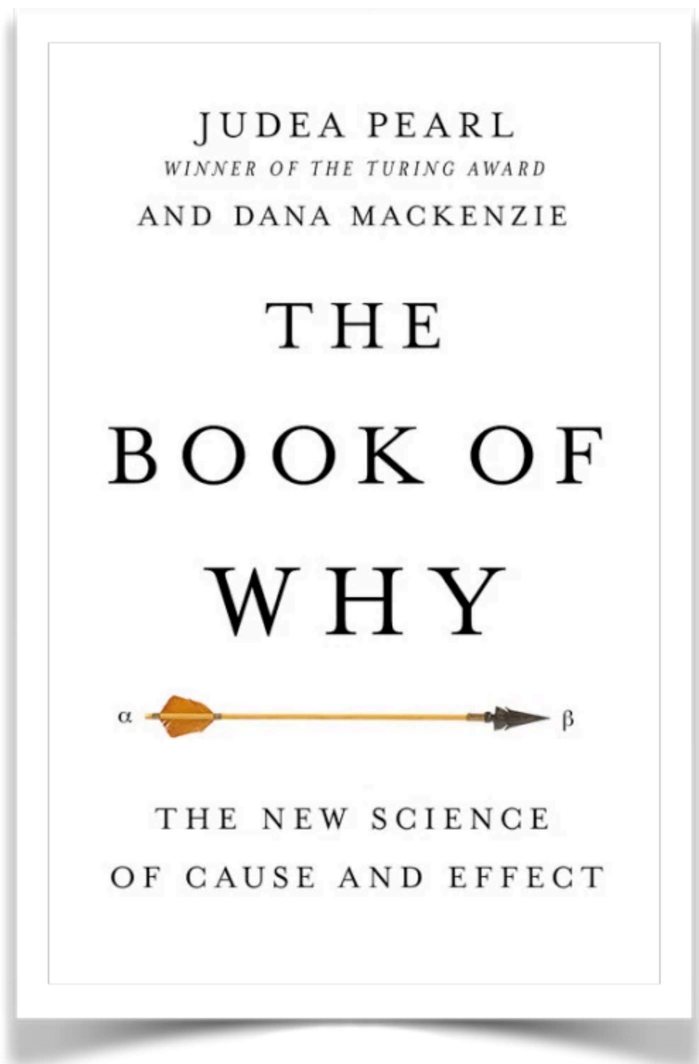
Causal modelling is a prime contribution to science

Judea Pearl received one of the highest scientific prizes for his ideas

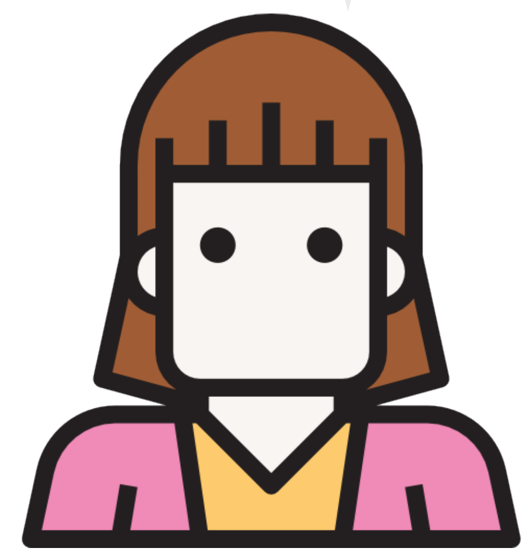
Maybe that's just gibberish of another mad professor!



Don



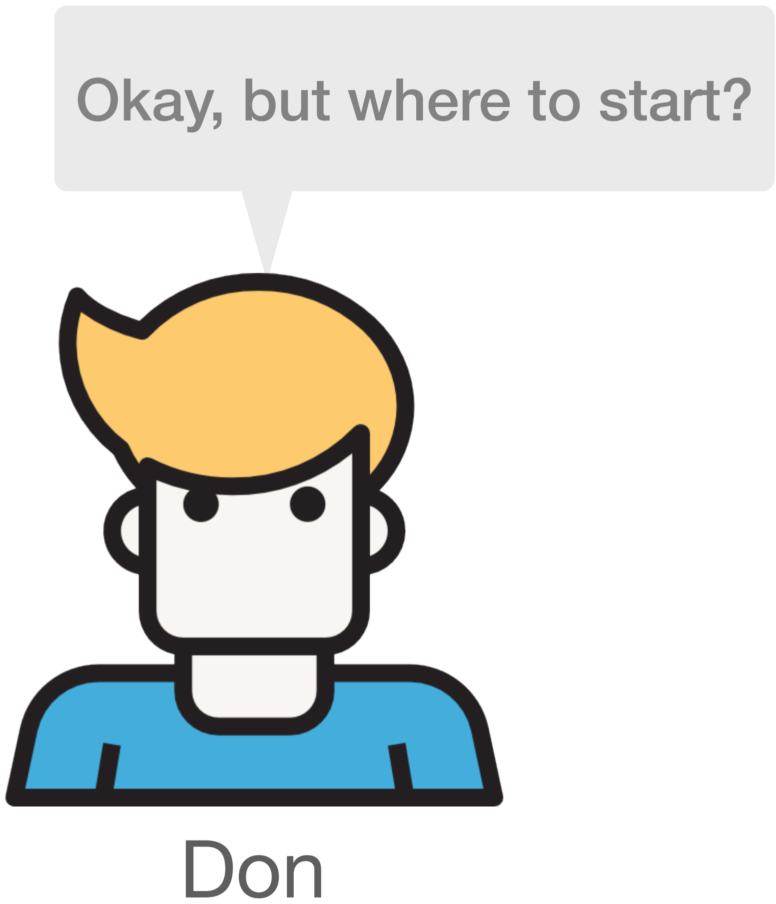
Well, one of the main authors, Judea Pearl, won the Turing Award for his ideas as presented here.



Angi

Literature to get your hands dirty

Introductory literature to causal modelling in data analysis



- 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

Literatur

Corvetti, C. (2006). Saratoga Houses. https://rdrr.io/cran/mosaicData/

Dablander, F. (2020). An Introduction to Causal Inference [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/b3fkw

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