Causal Inference: Why We Should and How We Can Teach it in Introductory Courses

Karsten Lübke

CAUSE Webinar - June 9th, 2020

EDC Oceans of Data Institute (2016): The data-literate individual understands, explains, and documents the utility and limitations of data by becoming a critical consumer of data, controlling his/her personal data trail, finding meaning in data, and taking action based on data. The data-literate individual can identify, collect, evaluate, analyze, interpret, present, and protect data.

Some lessons you may already know or learn

To take the best action or causal conclusion based on multivariate (observational) data analysis:

- Data is not just there - it has a generating process and we should care about this.
- Confounding and bias can be serious issues for causal inference.
- Adjusting or not adjusting: Both can be bad ideas for causal inference.
- Structural causal models and directed acyclic graphs can help to build a bridge between reality, theory and data.
- Quantitative model checks may not reveal which model is best for causal inference (*only claimed, but true nevertheless*).

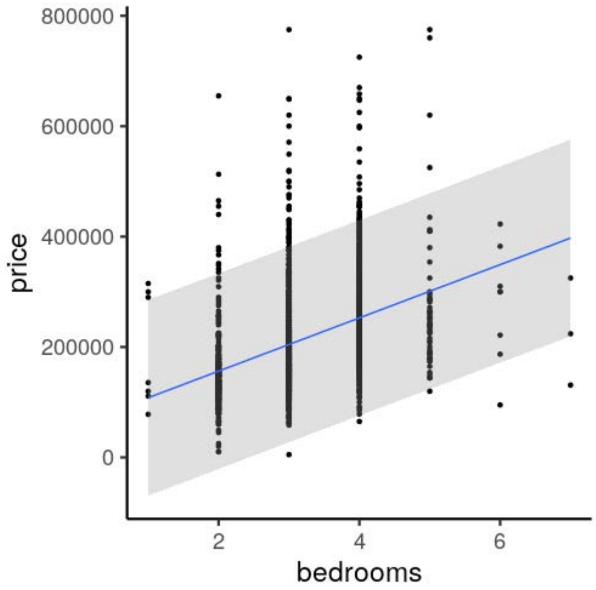
👉 For Intro Courses: Directed acyclic graphs may help to develop a framework to think about the data generating process.

Real data example

Saratoga Houses

Data on houses in Saratoga County, New York, USA in 2006. Analysis in `R`:

```
# Load package and read in data
library(mosaic); data(SaratogaHouses)
# Scatterplot
gf_point(price ~ bedrooms, data = SaratogaHouses) %>%
  gf_lm(interval = "prediction")
```



Idea: De Veaux (2019). *Data Science for All*

Modelling value of my 2-bedroom house

Linear Model: $\text{price}_i = \beta_0 + \beta_{\text{bedrooms}} \times \text{bedrooms}_i + \epsilon_i$:

```
# Linear Regression
my.model <- lm(price ~ bedrooms, data = SaratogaHouses); my.model
```

```
##
## Call:
## lm(formula = price ~ bedrooms, data = SaratogaHouses)
##
## Coefficients:
## (Intercept)     bedrooms
##       59863        48218
```

So: $\hat{\beta}_{\text{bedrooms}} = 48217.81$

```
# My house: 2 bedrooms; Point prediction
My.House <- data.frame(bedrooms = 2); predict(my.model, newdata = My.House)
```

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

$$\widehat{\text{price}}^{|\text{bedrooms}=2} \approx 156299$$

Turn data into money

Split one bedroom into three!

```
# My rebuilt house: now 4 bedrooms
My.NewHouse <- data.frame(bedrooms = 4)
# My money (?)
predict(my.model, newdata = My.NewHouse) - predict(my.model, newdata = My.House)
```

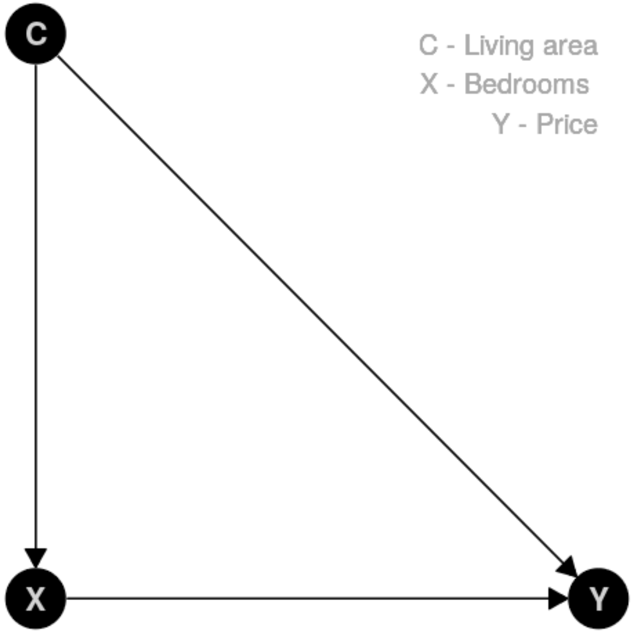
```
##        1
## 96435.62
```

So:

$$\widehat{\text{price}}^{|\text{bedrooms}=4} - \widehat{\text{price}}^{|\text{bedrooms}=2} = (4-2) \times \hat{\beta}_{\text{bedrooms}} = 96435.62$$

Causal Model (simplified)

The number of bedrooms depends on the house size - as well as the price (**confounding**/ **lurking** variable):



Omitted-variable bias

Ok, let's **adjust** for `livingArea`:

```
my.adj.model <- lm(price ~ bedrooms + livingArea, data = SaratogaHouses); my.adj.model
```

```
##
## Call:
## lm(formula = price ~ bedrooms + livingArea, data = SaratogaHouses)
##
## Coefficients:
## (Intercept)     bedrooms   livingArea
##     36667.9     -14196.8        125.4
```

Now: $\hat{\beta}_{\text{bedrooms}} = -14196.77$ (instead of $\hat{\beta}_{\text{bedrooms}} = 48217.81$ unadjusted for `livingArea`). So: price falls instead of rises if I split a bedroom. (**Simpson's Paradox**)

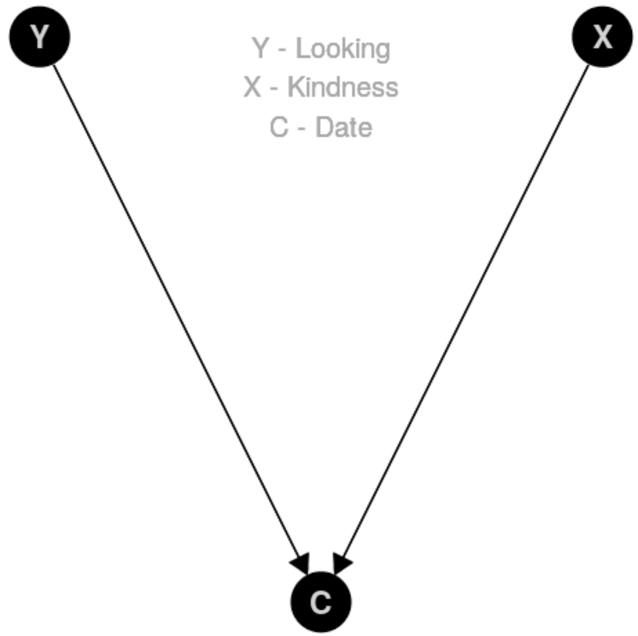
No problem, just use a fancy-machine-learning-method with all variables? (And be aware of bias-variance trade off.)

Unfortunately quantitative measures like e.g. cross-validated mean squared error **not always** tell you which model is best for causal inference.

Simulated example

Dating

Assume you date someone because he is good looking or because he is kind. Moreover assume that looking and kindness are independent.



Really?

Make three bedrooms out of one and the value of my house goes up by ≈ 100.000 Dollar?

via GIPHY

Data generating process

$$X = U_X, \quad U_X \sim \mathcal{N}(0,\, 1),$$

$$Y = U_Y, \quad U_Y \sim \mathcal{N}(0,\, 1),$$

$$\widetilde{C} = \begin{cases} 1 & \text{, if } \{X > 1 \vee Y > 1\} \\ 0 & \text{, else} \end{cases},$$

$$C = (1 - U_C) \cdot \widetilde{C} + U_C \cdot (1 - \widetilde{C}), \quad U_C \sim \mathcal{B}(0.05).$$

where $\mathcal{N}(\mu,\, \sigma)$ stands for Normal distribution and $\mathcal{B}(\pi)$ for the Bernoulli distribution.

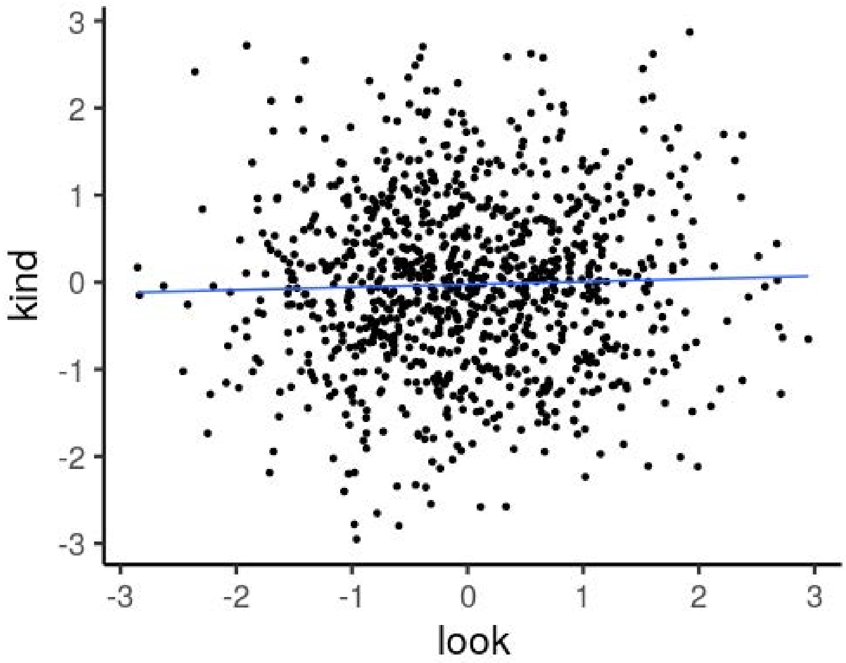
In R:

```
set.seed(1896)

kind<- rnorm(1000)
look <- rnorm(1000)
dating <- ((kind > 1) | (look > 1))
luck <- rbinom(1000, size = 1, prob = 0.05)
dating <- (1 - luck) * dating + luck * (1 - dating)
```

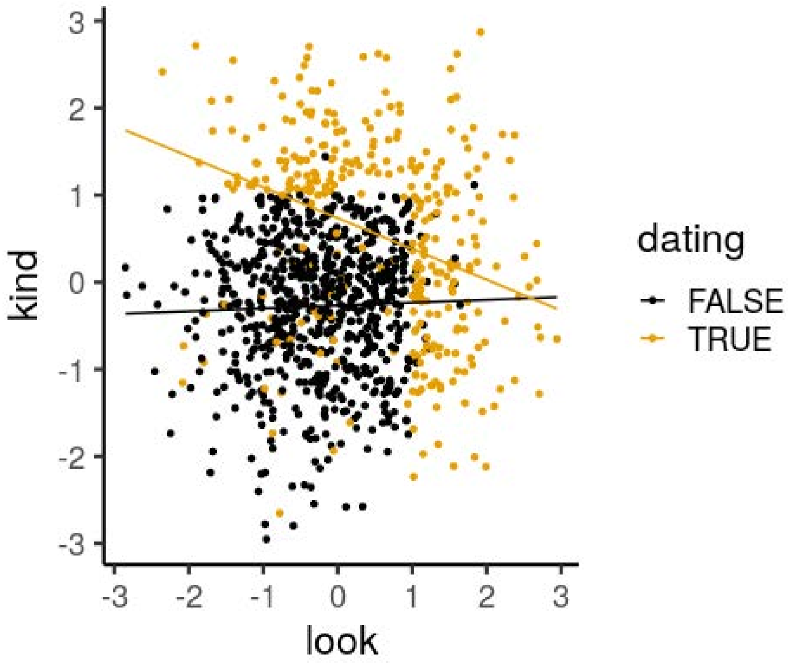
Modelling: Marginal

Modelling of `kind` by `look`:



Modelling: Conditional

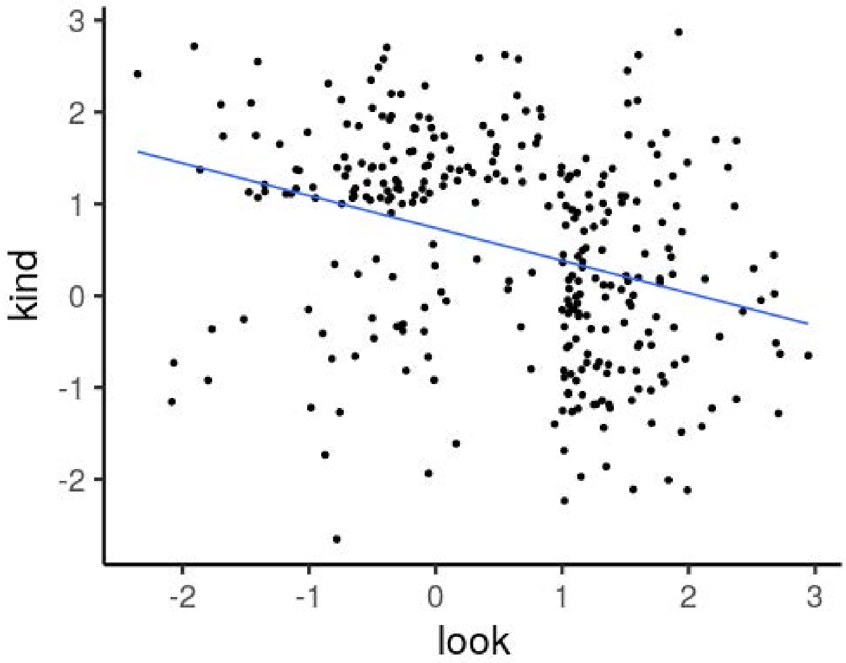
Modelling of `kind` by `look`, adjusted for `dating`:



- Adjusted for the common effect (dating) there is an association between the independent causes good looking and kindness (**Berkson's Paradox**).
- Formally: $\texttt{kind} \perp\!\!\!\perp \texttt{look}$ *but* $\texttt{kind} \not\perp\!\!\!\perp \texttt{look} | \texttt{dating}$

Selection/ Collider Bias

Modelling of `kind` by `look`, selected by `dating`:



- There is also an association between the independent causes good looking and kindness if your data consists only of those who you dated - if good looking was not the reason, it must have been kindness (or luck).

Excursion: Our looks

Kind *and* good looking people who contributed to this work:

Matthias Gehrke, Jörg Horst, Gero Szepannek, Bianca Krol, Sebastian Sauer

Your presenter:

Some more words about data science and causal inference

Data science tasks

via GIPHY

- Shmueli (2010) asked: *To Explain or to Predict?*
- Hernán et al. (2019) distinguished:
    - **Description**: "How can women aged 60–80 years with stroke history be partitioned in classes defined by their characteristics?"
    - **Prediction**: "What is the probability of having a stroke next year for women with certain characteristics?"
    - **Causal inference**: "Will starting a statin reduce, on average, the risk of stroke in women with certain characteristics?"

Levels of causal inference

Pearl (2019) establishes a three-level hierarchy:

- **Association**: $P(y|x)$: Seeing: *what is?*, i.e., the probability of $Y = y$ given that we observe $X = x$.
- **Intervention**: $P(y|do(x))$: Manipulation: *what if?*, i.e., the probability of $Y = y$ given that we intervene and set the value of X to x.
- **Counterfactuals**: $P(y_x|x', y')$: Imagining: *what if I had acted differently?*, i.e., the probability of $Y = y$ if X had been x given that we actually observed x', y'.

Other approaches to causal inference are e.g. within potential outcome framework, instrumental variables, regression discontinuity designs, Granger, natural experiments, ...

One page of theory

- $X \to Y: \quad Y = f(X, U_Y)$ with some function $f(\cdot)$ and some exogenous U.
- The value of Y depends on X - but the value of X **not** on Y.
- Causally there is no inverse function $f^{-1}(\cdot)$. My weight growths with my height but unfortunately my height not with my weight.

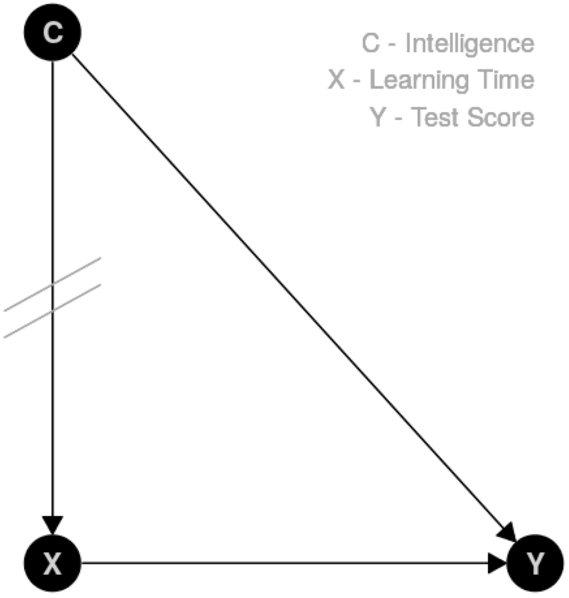
Path	$X \to C \to Y$	$X \leftarrow C \to Y$	$X \to C \leftarrow Y$
Name	Chain	Fork	Collider
Association X to Y	Causal	Non-causal	None
Role of C	Mediator	Cause	Effect
Adjusting C	Blocks causal path	Blocks non-causal path	Opens biasing path

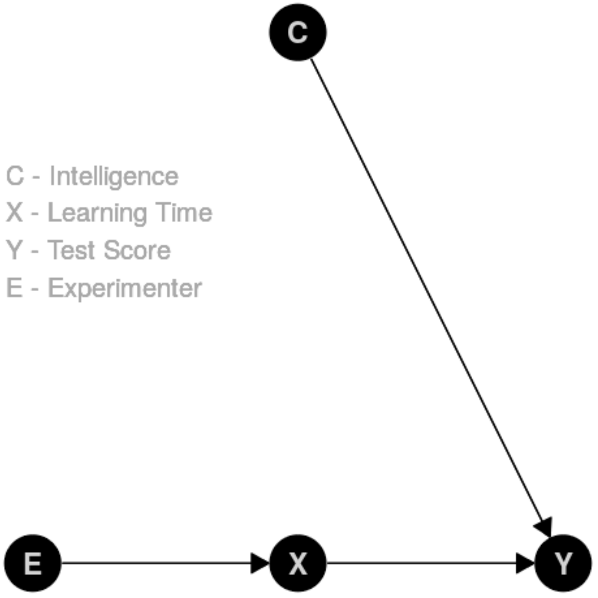
Idea: To estimate the change in y if x is changed: Block non-causal paths, open causal paths and don't open a biasing path.

BTW: DAGitty is a nice browser-based environment for creating, editing, and analyzing causal diagrams (directed acyclic graphs).

Randomized controlled trial

Within a randomized controlled trial observation i is randomly assigend to x_i: $do(x_i)$. The connections to the *parents* of X is cut:





Teaching Causal Inference

Witmer (2020): The scientific community would benefit greatly from a better understanding of causal inference - and "better" is quite a low bar, given how little the tools of causal reasoning have been used over the years. But statisticians have stood in the way, insisting that cause-and-effect conclusions can only be drawn from randomized experiments and delighting in telling stories about confounded effects that arise when analyzing observational data, all while repeating the mantra that correlation is not causation. In so doing, we statisticians congratulate ourselves too much, while turning students away from asking and answering questions of genuine interest.

Changing curriculum

Cobb (2015):

> Mere Renovation is Too Little Too Late: We Need to Rethink our Undergraduate Curriculum from the Ground Up

Influenced by GAISE (2016), but also Kaplan (2018) and Schield (2018):

- Wild and Pfannkuch (1999). Statistical Thinking in Empirical Enquiry
- Blog Lindeløv (2019). Common statistical tests are linear models (or: how to teach stats)
- Pruim, Kaplan and Horton (2017). The mosaic Package: Helping Students to 'Think with Data' Using R
- Rossman and Chance (2014). Using simulation-based inference for learning introductory statistics (cf. Blog Downey (2018)).

Together with Reproducible Analysis, Quizzes, Fun Elements, shiny and learnr apps.

Inspired by e.g. the -books Open Intro (ISRS), ModernDive, Statistical Modeling (2e) and Data 8.

Asking questions

- Is it a good idea to show oversimplified examples in class? Real causal inference is much harder.
- Are we overstraining our students?
- Simulated data, really?
- What about the topics ommited: Is causal inference making up the opportunity costs?
- What happens if our students change university and have learned different topics than in a consensus curriculum?
- Is it a good idea to teach something that most of us have not learned as a student?
- On the other hand, are we answering the important questions for data literacy in the consensus curriculum?

Outro

Want more?

Some References:

- Cummiskey, K., Adams, B., Pleuss, J., Turner, D., Clark, N., & Watts, K. (2020). Causal Inference in Introductory Statistics Courses. Journal of Statistics Education
- Dablander, F. (2019). An introduction to Causal inference (Blog)
- 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.
- Elwert, F. (2013). Graphical causal models. In: Handbook of causal analysis for social research (S. 245-273). Springer, Dordrecht.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.
- Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference: foundations and learning algorithms. MIT press.

Also:

Several R packages exists, e.g. `ggdag`.

Own Work and contact

- 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. (gives more examples and details)

- Lübke, K. & Gehrke, M. (2020). *Now is the Time for Causal Inference in Introductory Statistics*, Proceedings IASE 2020 Roundtable New Skills in the Changing World of Statistics Education (accepted). (will give more background on the teaching aspects)

- -material for this webinar: https://github.com/luebby/CAUSE-Webinar

- : karsten.luebke@fom.de
- : @luebby42
- : @luebby

Thank you!

Recap and Outlook

Causal Inference: Why We Should and How We Can Teach it in Introductory Courses

👉 Directed acyclic graphs may help to develop a framework to think about the data generating process (in a world full of multivariate observational data).

Pearl (2019): Seven Sparks of Causal Inference

1. Encoding causal information in transparent and testable way
2. Predicting the effects of actions and policies
3. Computing counterfactuals and finding causes of effects
4. Computing direct and indirect effects (Mediation)
5. Integrating data from diverse sources
6. Recovering from missing data
7. Discovering causal relations from data