DS 102: Data, Inference, and Decisions

Lecture 13

Michael Jordan
University of California, Berkeley

Correlations and Causes

- You've (hopefully) heard throughout your education to not make the mistake of thinking that a correlation implies a causal relationship
  - unfortunately, far too many people haven't had your education
  - you'll find the mistake made very frequently (e.g.) in journalism (but not only in journalism)
- In this lecture, we'll invent a few spurious inferences of causality just to remind ourselves of the issue
- And then we'll tackle the harder (senior-level!) question: can we actually ever infer causality? If so, how?

Mistaking Correlation for Cause

- Let's take a moment to invent some examples
- I'll ask each of you to think up an example and write it in the chat box
- To help seed the process, here's one that I invented:

A Martian arrives on planet Earth, and after a year of study announces the discovery of a correlation between the wearing of a coat and the common cold: people who wear a coat have a higher probability of having a cold than people who don't wear a coat. The Martian infers that the wearing of a coat causes the common cold.

Another Example (from Wikipedia)

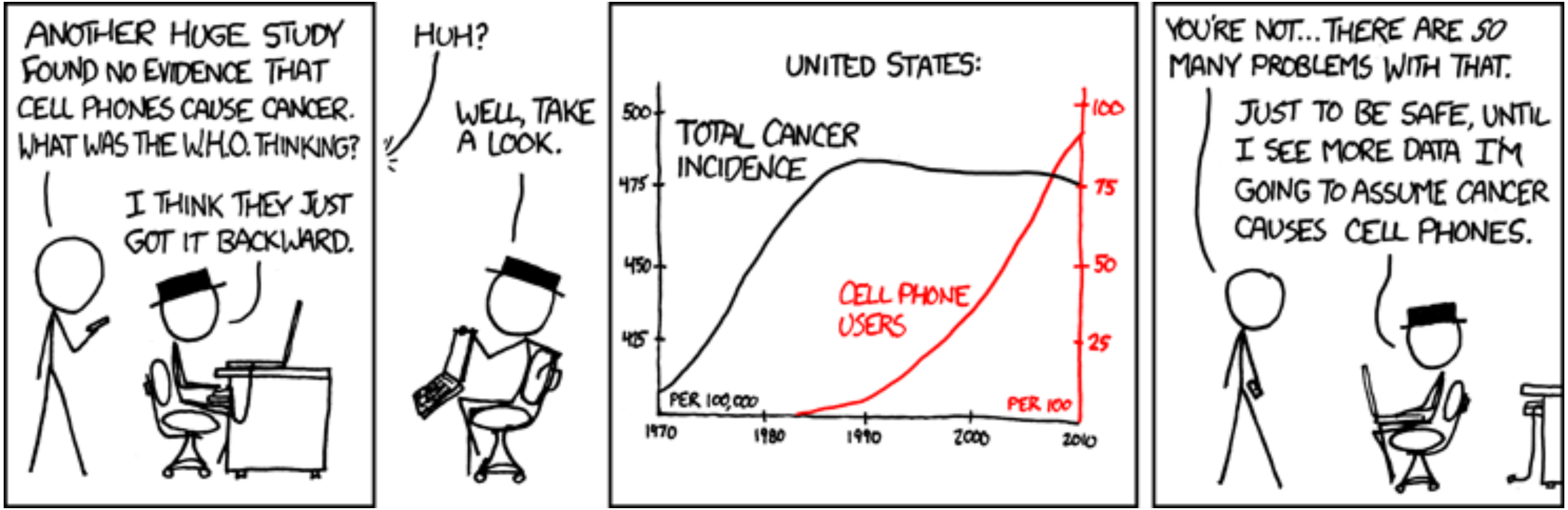
- *Europeans in the Middle Ages believed that lice were beneficial to your health, since there would rarely be any lice on sick people. The reasoning was that people got sick because the lice left.*

Another Example (from Wikipedia)

- *Europeans in the Middle Ages believed that lice were beneficial to your health, since there would rarely be any lice on sick people. The reasoning was that people got sick because the lice left.*

- *The real reason that the lice left is because they are extremely sensitive to body temperature.*

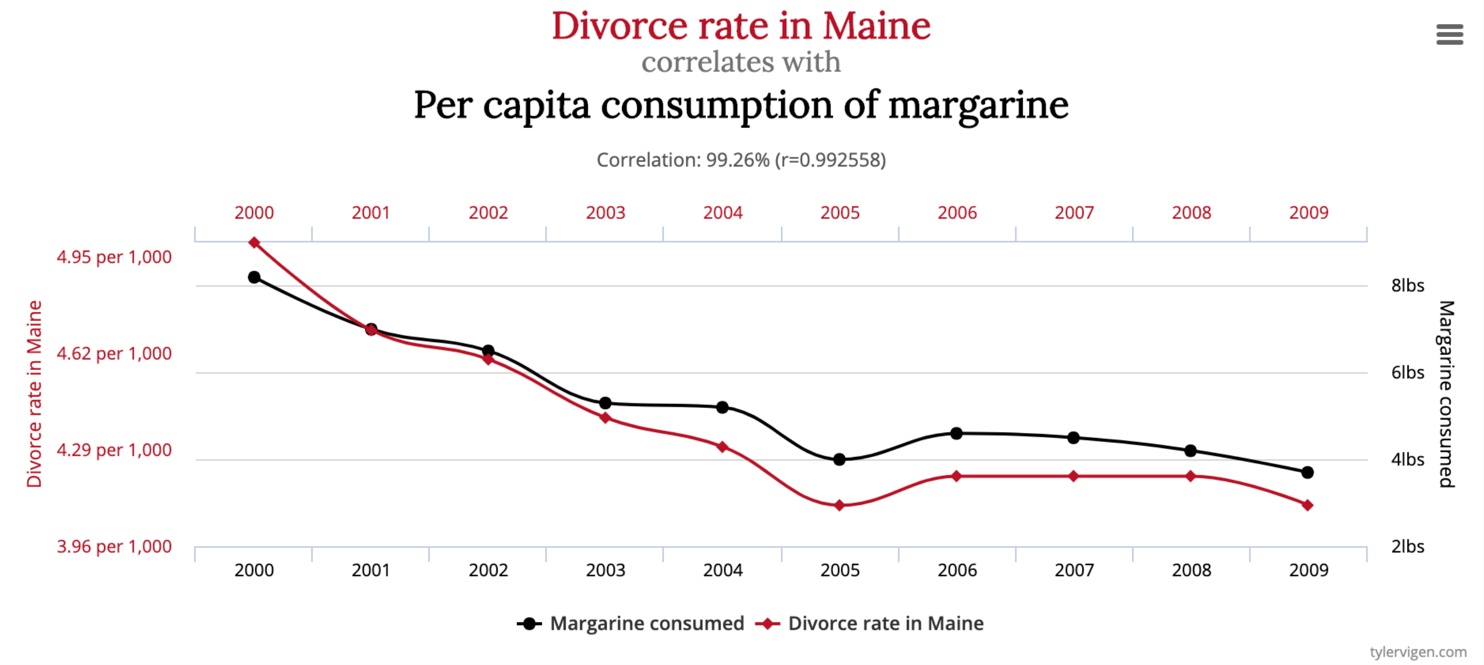
Another Example



Spurious Correlations

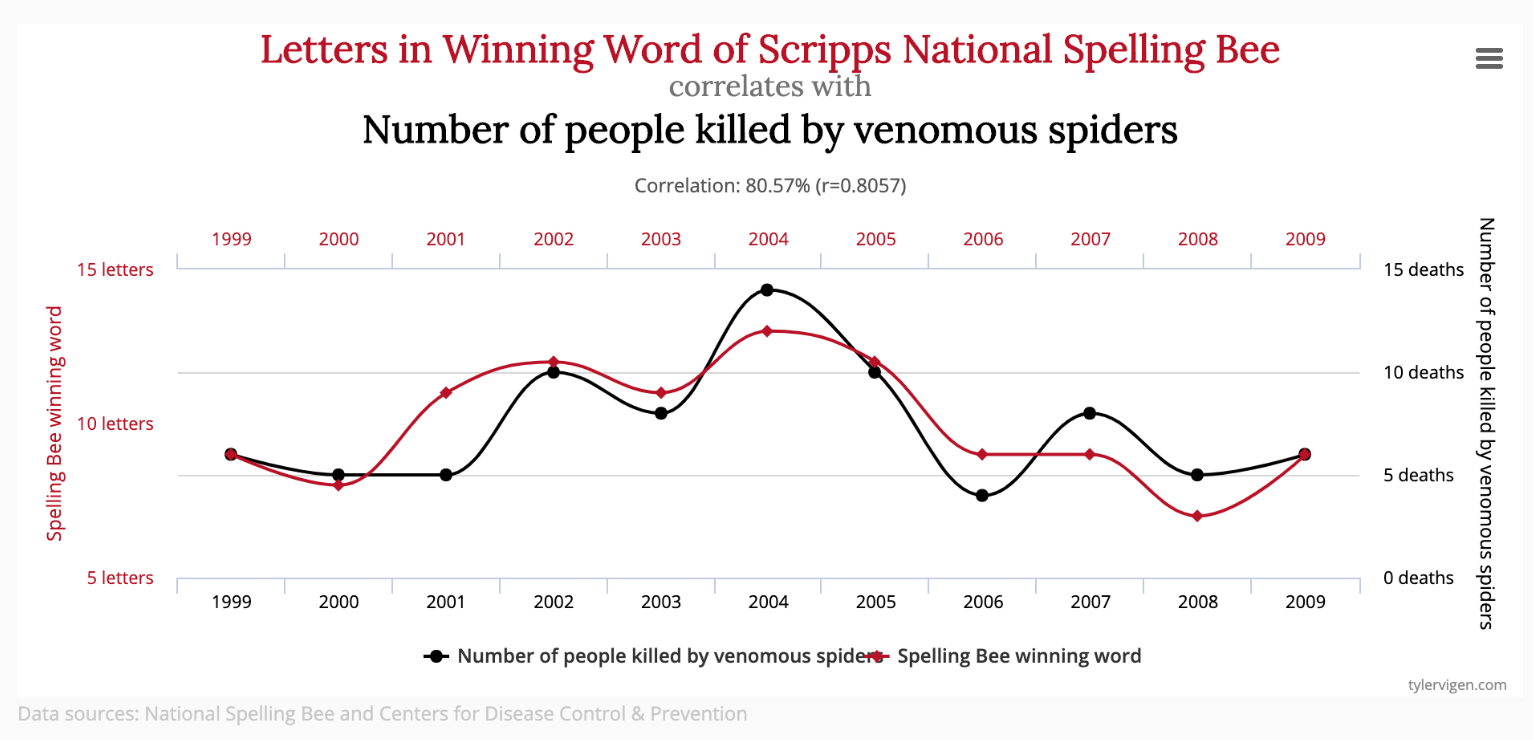
- Correlations found in empirical data need not reflect “true correlations”
- That doesn’t prevent humans from trying to find causal interpretations not only for real, but non-causal correlations, but also for spurious correlations…

Spurious Correlations (tylervigen.com)



Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

Spurious Correlations (tylervigen.com)



From Correlation to Cause

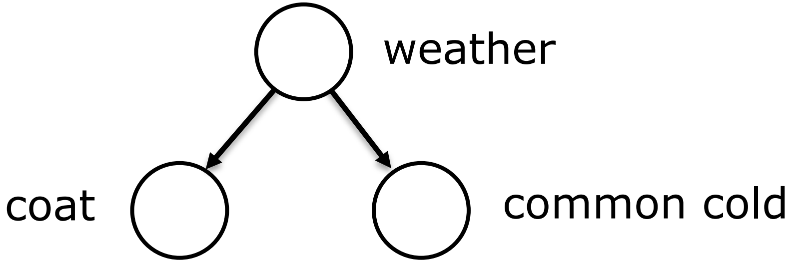
- It’s easy and fun to come up with such examples
- How do humans avoid making this mistake much of the time? How do they infer causes?
  - briefly, they often have “background knowledge”
  - but, particularly in new scientific fields, such knowledge is absent
  - also, humans have many biases and gaps that they are unaware of and mistake for actual knowledge

From Correlation to Cause

- It’s easy and fun to come up with such examples
- How do humans avoid making this mistake much of the time? How do they infer causes?
  - briefly, they often have “background knowledge”
  - but, particularly in new scientific fields, such knowledge is absent
  - also, humans have many biases and gaps that they are unaware of and mistake for actual knowledge
- One way to infer causality is to do a random experiment
  - for example, deciding whether a vaccine causes an immune response can be achieved by randomly assigning people to vaccine or control in a clinical trial
- But arguably humans (e.g., children) can infer causes by merely observing events in the world

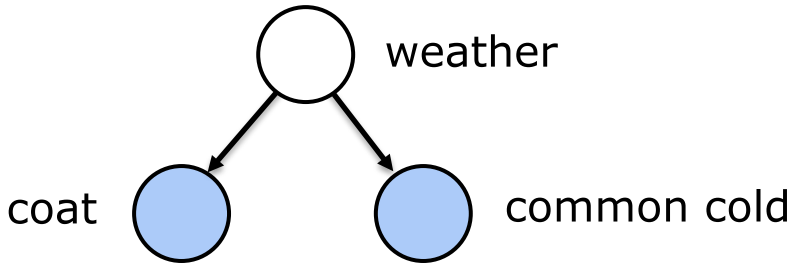
Latent Variables

- Many of the errors in mistaking correlation for cause come from the presence of an unmeasured “latent variable” that is the real cause and which affects both of the measured variables
  - in my example, the weather is the unmeasured variable
  - cold weather causes people to wear coats, and it also leads to more of the virus circulating, which causes the common cold



Latent Variables

- Marginalizing over the underlying latent variable induces an association between the observed variables:



Treatments, Controls, and Outcomes

- For simplicity, let's focus on assessing causality with respect to a binary variable, and refer to its two values as the treatment and the control

Treatments, Controls, and Outcomes

- For simplicity, let's focus on assessing causality with respect to a binary variable, and refer to its two values as the treatment and the control
- We use the integer variable $i \in \{1, \ldots, n\}$ to denote the experimental units

Treatments, Controls, and Outcomes

- For simplicity, let's focus on assessing causality with respect to a binary variable, and refer to its two values as the treatment and the control
- We use the integer variable $i \in \{1, \ldots, n\}$ to denote the experimental units
- To each experimental unit we apply either the treatment or the control, but not both
-

Treatments, Controls, and Outcomes

- For simplicity, let's focus on assessing causality with respect to a binary variable, and refer to its two values as the treatment and the control
- We use the integer variable $i \in \{1, \ldots, n\}$ to denote the experimental units
- To each experimental unit we apply either the treatment or the control, but not both
- We let $\{Z_i = 1\}$ denote the event that unit i is given the treatment, and let $\{Z_i = 0\}$ denote the event that unit i is given the control

Treatments, Controls, and Outcomes

- For simplicity, let's focus on assessing causality with respect to a binary variable, and refer to its two values as the treatment and the control
- We use the integer variable $i \in \{1, \ldots, n\}$ to denote the experimental units
- To each experimental unit we apply either the treatment or the control, but not both
- We let $\{Z_i = 1\}$ denote the event that unit i is given the treatment, and let $\{Z_i = 0\}$ denote the event that unit i is given the control
- The measured outcome associated with unit i is denoted Y_i

Potential Outcomes

- Inferential thinking often starts by considering what could have happened, and then trying to go backwards from data
- We define potential outcomes, $(Y_i(0), Y_i(1))$, as the hypothetical values that would be observed for unit i if that unit were assigned to the control or the treatment, respectively
- In reality, we only get to observe one of the potential outcomes, but conceptually both exist
- We can define the observed outcome as follows:

$$Y_i = \begin{cases} Y_i(0) & \text{if } Z_i = 0 \\ Y_i(1) & \text{if } Z_i = 1 \end{cases}$$

The Individual Causal Effect

- We define the individual causal effect to be:

$$\tau_i = Y_i(1) - Y_i(0)$$

  - e.g., does a person live longer if they're given the drug versus a control?

The Individual Causal Effect

- We define the individual causal effect to be:

$$\tau_i = Y_i(1) - Y_i(0)$$

  - e.g., does a person live longer if they're given the drug versus a control?

- Since we can observe either $Y_i(0)$ or $Y_i(1)$, but not both, it's pretty clear that with no further information we're not able to estimate the individual causal effect
- Instead, we aim to estimate the average causal effect (also known as the average treatment effect)

The Causal Inference Problem

Z_i	$Y_i(0)$	$Y_i(1)$
0	$\times$	?
$\vdots$	$\vdots$	$\vdots$
0	$\times$	?
1	?	$\times$
$\vdots$	$\vdots$	$\vdots$
1	?	$\times$

- How can we infer something about the difference between treatment and control when we never observe them together?

The Average Causal Effect

- We define the average causal effect to be:

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$$

  - where the expectation refers to independent and identical draws of the experimental units from an underlying population
  - and thus the value of τ is actually independent of i

- Can we estimate τ from data?