Share descriptions or links to stories about BAD causal inferences (or ones that might be bad) (please put your name next to things that you share)

Rahul Talari:

Bad Causal Inference - A causal relationship occurs when event 'A' influences event 'B' through 'flow of energy'. In some datasets, it is possible to conclude that one variable has direct influence on the other. A bad causal inference occurs when A's influence on B does not provide evidential information for B.

Example: A bank manager is concerned with the number of customers whose accounts are overdrawn. Half of the accounts that become overdrawn in one week are randomly selected and the manager telephones the customer to offer advice. Any difference between the mean account balances after two months of the overdrawn accounts that did and did not receive advice can be causally attributed to the phone calls.

Here the 'flow of energy' from overdrawn accounts → phone calls is not directly related. The overdrawn accounts can be related to stronger factors such as the economic status of the customer.

Name: Jiaqi
Bad Causal Inference:
Wearing helmets cause more bicycle-related accidents.
Observation 1: A nationwide study indicates that ten years ago, 35% of all bicyclists reported wearing helmets, whereas ten years later the number is 80%.
Observation 2: Meanwhile, another study show that during the same ten-year period, the number of bicycle-related accidents has increased 200 percent.
Bad Conclusion: Wearing helmets cause more bicycle-related accidents.
Comment: Bicycle-related accidents may cause by many factors, for example, there are more cars on the road, etc. There is no evident that not wearing helmets will reduce the bicycle-related accidents. (I pick this story from the GRE argument pool).

Name: Adam
Bad Causal Inference:
Everything from this site - http://www.tylervigen.com/spurious-correlations
The site does not claim that these are causal inferences, but are other things like the "ice cream case" where things are highly correlated and people could easily make bad inferences.
Examples:
# of people who die in swimming pools correlates with number of films that Nicolas Cage appeared in that year.

Divorce rate in Maine correlates highly with per capita consumption of margarine.

Name: Sriram (aka ram)
Bad causal inference:
observation 1 : A group of 8 friends went out for dinner.
observation 2 : 5 of them ended up having a stomach ache on the next day.
observation 3 : all 5 of them had dish which had cheese.
observation 4 : all 5 of them were lactose intolerant.
observation 5 : Cheese has lactose.
A bad inference will be to say having dishes with cheese causes stomach ache. Because the reason for stomach ache came from the fact that all of them were lactose intolerant and cheese has lactose.

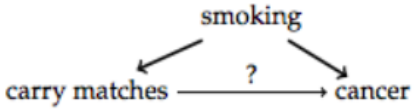
Nikhil Kini
As long as popular media exists and tries to popularize scientific results, there will always be great examples of bad causal inference. The most popular one might be - Listening to Mozart increases your IQ. This one is so popular, it even has its own Wikipedia page.
The obvious problem is certainly the jump from what the paper claims to what the people understood. Even the paper itself is actually not talking about the "increase in a person's IQ", it merely compares the scores on a spatio-temporal reasoning test of 3 groups of people, where the difference between the groups was which relaxation technique was adopted 10 minutes before administering the test. Those who relaxed by listening to Mozart scored 8-9 IQ points higher. There was no "increase in IQ". However, there were experiments performed later which actually tried to concretize the theory that Mozart can indeed *increase* one's IQ. The findings remain controversial/contested.

Here is a good resource on articles in popular media about causation vs correlation and misrepresented research.

Diego Rodriguez
Here is an example where correlation is confused by causality:
Headline: "Dogs Walked by Men Are More Aggressive"
Extract from the article: "The study, accepted for publication in the journal Applied Animal Behavior Science, surprisingly found that the sex of the owner had the biggest effect on whether or not the dog would threaten or bite another dog"
Source: http://news.discovery.com/animals/dog-walking-behavior-111103.html

Paris Miri

Carrying matches in one's pocket and developing lung cancer (example of a **confounded** association between two variables, which is mediated via the biasing path (bold)).

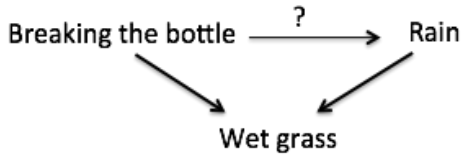


If we were to perform an association study on the relationship between carrying matches in one's pocket and developing lung cancer, we would probably find a correlation between these two variables. However, as the above diagram indicates, this correlation would not imply that carrying matches in your pocket causes lung cancer: Smokers are more likely to carry matches in their pockets, and also more likely to develop lung cancer.

This is Pearl's favorite example in most of his talks:

1) If the grass is wet then it rained.
2) If we break the bottle, the grass will get wet.

Therefore we can infer that if we break this bottle then it rained.



Pigi

I will also give a bad causal inference example for the case we have two correlated variables but there is no causation between those.

Example: Please see the graph here: http://www.fastcodesign.com/3030529/infographic-of-the-day/hilarious-graphs-prove-that-correlation-isnt-causation#3 shows that the amount of money that the US government is spending on science, space, and technology is correlated to the suicides by hanging strangulation and suffocation. So a false causal inference that we could make is that the increase in the amount spent by the US government on science, space, and technology increases the suicides by hanging strangulation and suffocation.

I also found this interesting article http://jeremyfreese.blogspot.com/2005/06/still-another-example-of-bad-causal.html which states that "**research showing that ex-prisoners who vote are less likely to end up back in prison.**" However, I wasn't able to find the research study mentioned in the article in order to reach a safe conclusion.

Natalia

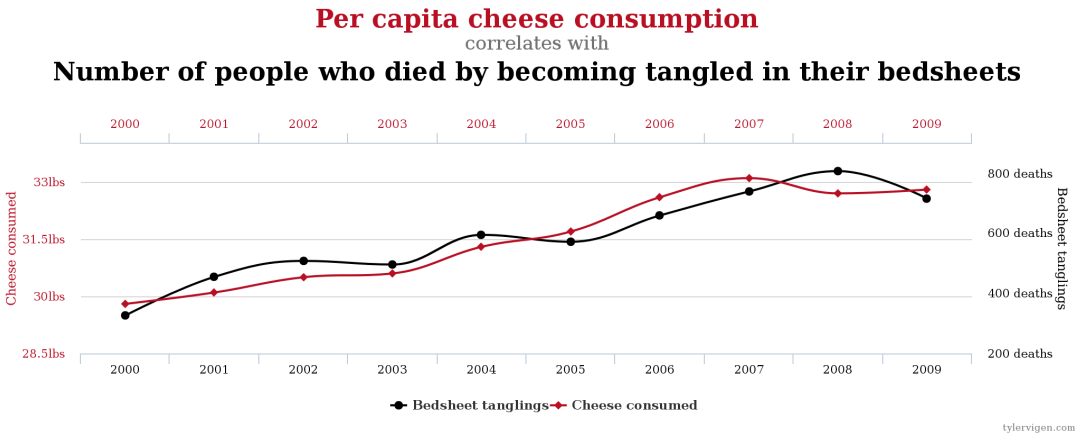
Correlations among different unconnected facts can give rise to confusion among causality and correlation, when we are talking purely about correlation by casualty.

Example 1: The total revenue generated by arcades correlates with the number of Computer Science PhD degrees awarded in the U.S. Source: http://tylervigen.com/spurious-correlations

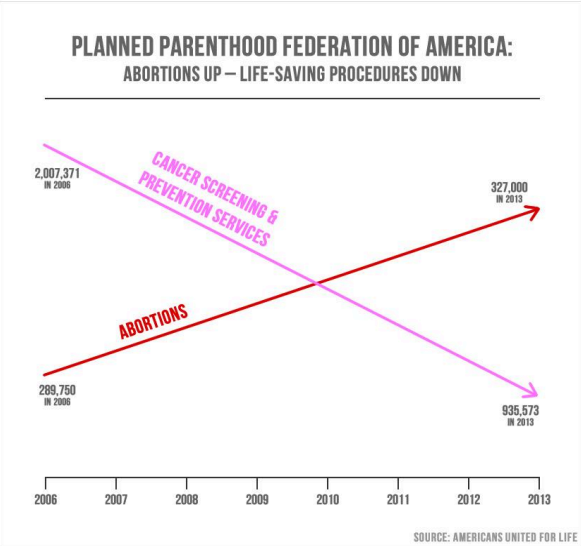
Example 2: The amount of chocolate intake per capita correlates with the amount of Nobel Laureates. This can read in news as the causation, but correlation does not imply causation! http://knowingneurons.com/2013/06/17/correlation-vs-causation-chocolate-consumption-and-nobel-laureates/

Matt H.

I agree with the authors who posted Tyler Vigen's: Spurious Correlations site for examples of quirky and ridiculous correlations from government data. My personal favorite:



As some of you may have seen, a bad example of causality caught a lot of media attention in the past week. An elected US Congressman used this chart to argue against funding for the organization, Planned Parenthood.



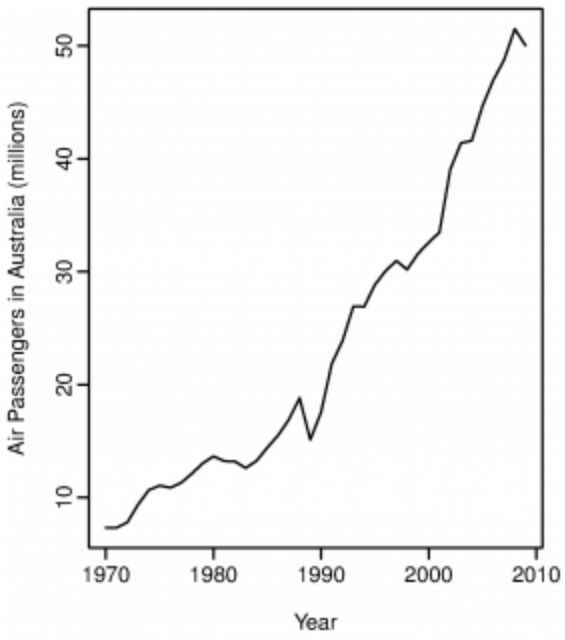
This example really demonstrates how even loose correlation can be used to mislead people about causal effects, and also how you can draw incorrect conclusions by focusing too narrowly on a small set of variables.

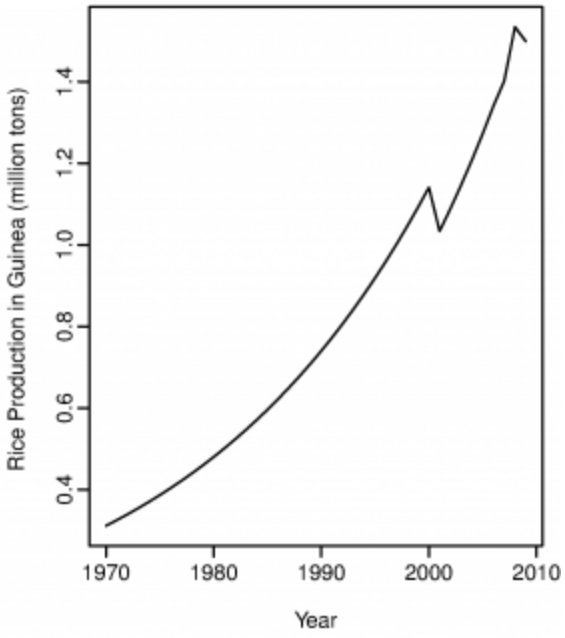
Johnathan

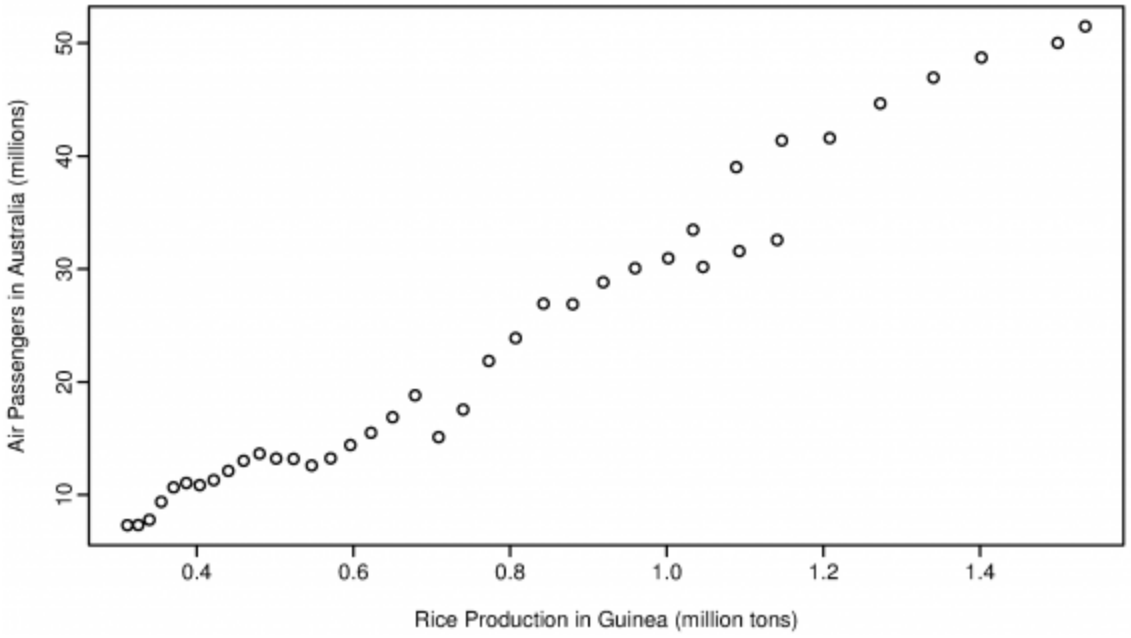
When in doubt, just go with a list:
http://www.csus.edu/indiv/m/mayesgr/phl4/Handouts/phl4causalfallacies.htm
This outlines causal fallacies and provides examples for each one. The list covers a pretty good range, although I'm sure we could find more. Summed up:
-- *Post Hoc*: Asserting A is the cause of B because A happened before B
-- *Reversing Cause and Effect*: A and B often occur together, therefore either A causes B or B causes A
-- *Negating a Common Cause*: Assuming that A causes B or B causes A and negating the third possibility that C causes A and B.
-- *Causal Determinism*: If A causes B, than when A occurs, B *must* occur.

Keshav

This one I saw over the summer and had been searching for. Classic example of how causal reasoning can be completely wrong when based on such a correlation. Rice production in Guinea vs Air Passengers in Australia. [https://www.otexts.org/fpp/4/8]


Air Passengers in Australia (millions)
10
20
30
40
50
1970
1980
1990
2000
2010
Year


Rice Production in Guinea (million tons)
0.4
0.6
0.8
1.0
1.2
1.4
1970
1980
1990
2000
2010
Year

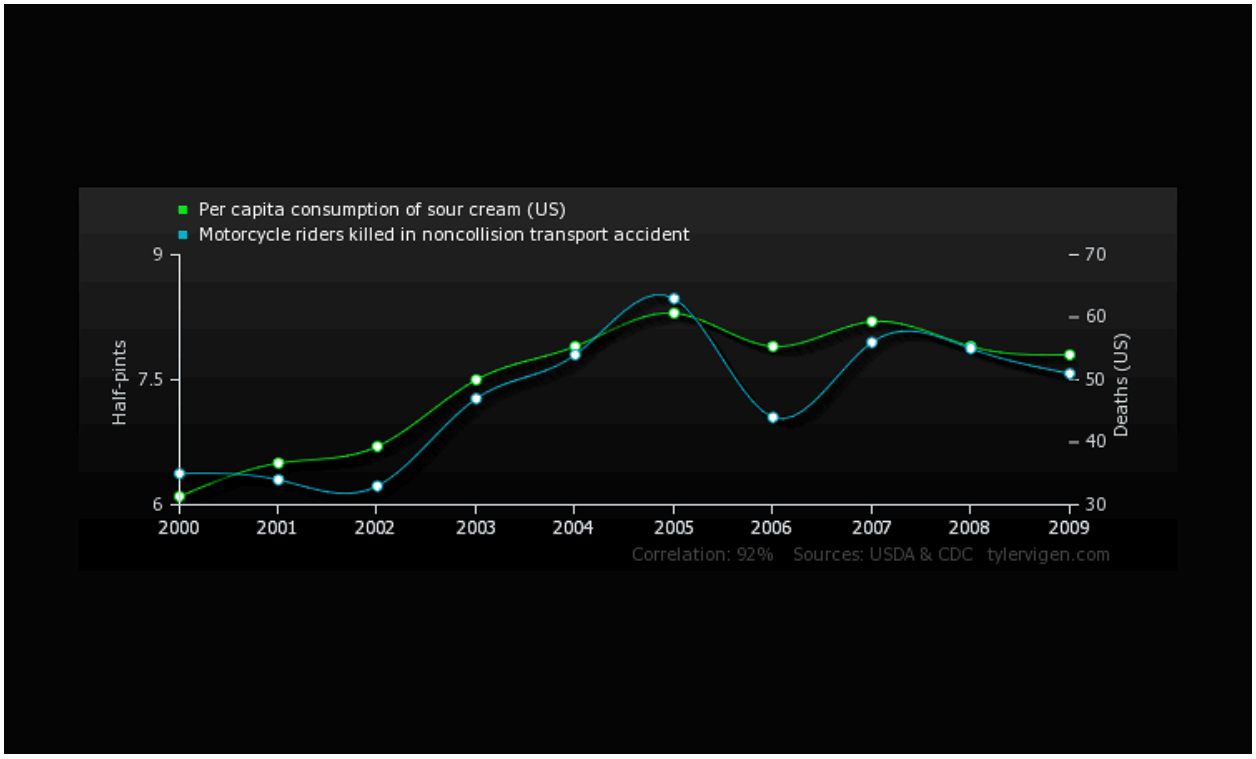

Air Passengers in Australia (millions)
10
20
30
40
50
0.4
0.6
0.8
1.0
1.2
1.4
Rice Production in Guinea (million tons)

Shachi

1) I had come across a study which said that milk consumption is correlated with hip fractures in women. But the research here : http://www.ncbi.nlm.nih.gov/pubmed/20949604 indicates that there was no overall association between milk consumption and hip fracture risk in women.

2) Came across this one : shoe size vs reading ability : This is an example of correlation but it is a bad causal inference.

3) Found this on http://www.fastcodesign.com/3030529/infographic-of-the-day/hilarious-graphs-prove-that-correlation-isnt-causation. The graph below shows that consuming sour cream and #motorcycle riders killed in noncollision transport accident are highly correlated. But this should not be mistaken for causation.



I also really like the one Pigi already mentioned : US spending on technology, science being highly correlated with suicides caused by suffocation/hanging, which is clearly a bad causation example.

Dhanya

Bad causal inference: (1) attributing cold weather to the actual viral infection that we know as the common cold. The latent causal factor is often that in cooler weather, people stay in indoors where infection is much more likely to spread quickly.

Ryan

I went to interview at Indiana University some years ago and there a professor that was very excited to share with me how they found that Twitter Mood predicts the Stock market. Recently I returned to check if the results stayed valid only to find out that it was more so a result of data mining bias that any actual effect. This outlines a good rebutal to the original paper:
http://sellthenews.tumblr.com/post/21067996377/noitdoesnot
While the original paper is careful to outline any causality as they do sound very surprised in the finding, the statistical methods they use are meant for finding causality and do return a positive result.

Morteza
There have been many (sometimes funny) cases that I have heard around, but the one that I experienced personally, was when I ran a pilot study during my MS in which it appeared that younger participants are better in playing games socially. It was tempting to say that the younger you are, the more willing you are to engage more fully in a social game (having seen kids as an extreme, it *could* make sense). However, after I investigated it more closely, turned out that my participants (who were coming from a pool in which they were offered credits for a course in social sciences), were coming from two different courses which they normally take in freshman year and senior year. After talking to a social science professor who was more familiar with the issue, we concluded that the freshman students are more "careful" in participating in studies, therefore they try to perform better and (play the game better), but the senior students perhaps care less (maybe because of having already established their GPA, etc.) and therefore perform somewhat poorly. This conclusion from the pilot study helped me to change the participant pool for the main study.

Sabina
I recently read this paper, "The Parable of Google Flu: Traps in Big Data Analysis", and I wanted to bring it up because there is a weird interaction that I haven't seen before this paper. Google has this flu prediction tool, and recently it was called out for actually being pretty bad at predicting flu outbreaks. One of the problems with it, that is related to causation is that, there was a possible effect because Google was overpredicting the flu, so the ads and the search results were tailored for people that were predicted to have the flu, which possibly influenced them to then search for more flu terms, so you have a weird self fulfilling prophecy of people thinking they have(and possibly being led to believe they have) the flu searching for flu symptoms which feeds back into the predictions, and so you the correlations aren't between the things you care about. Anyway it's maybe not super related but I wanted to bring it up.

Another example is something that I feel like I encounter a lot in conversation, and encountered a few months ago in a particularly uncomfortable conversation, which is the argument that people are poor because they are stupid. So that is a bad example of a causal inference. Maybe the truth for that example is that there are worse schools in lower income areas, and maybe no schools, and that tests for intelligence are biased towards people with educations.