

Experimental Design

What do you want to test? (explore and define the topic)

What is your hypothesis? (formulate a specific question to which you want to know the answer)

How does this fit into what you and others already know?

Independent and dependent variables: what are you going to change deliberately, and what are you going to measure? ONE INDEPENDENT VARIABLE AT A TIME FOR CLARITY

What are you going to hold constant? How are you going to hold it constant?

How are you going to do it?

- *experiment should be repeatable (by you and by others)
- *feasible in the time and \$\$ you have available?
- *should produce results that will test your hypothesis!

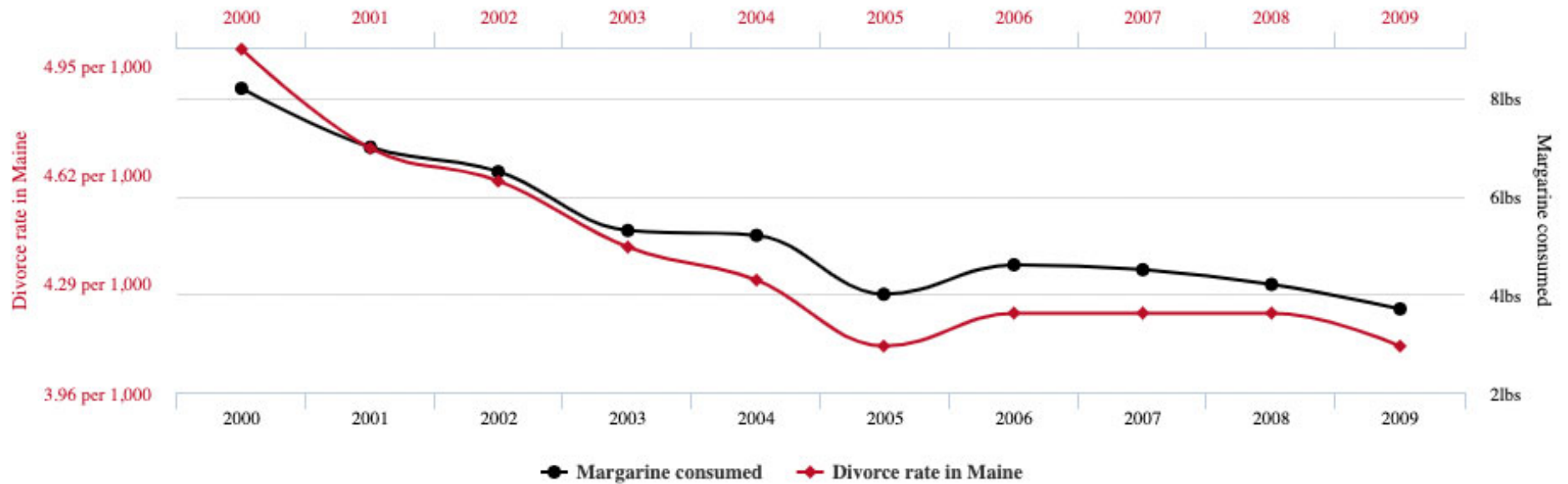
Is this even a good idea? (ethical implications)

Pitfalls in conducting an experiment?

- *method introduces too much error to test what you wanted to?
- *unintended variables not held constant (environment)
- *confirmation bias: give your data room to tell their actual story, not the one you expected them to tell!
- *correlation vs. causation (overinterpreting? P-hacking?)

Divorce rate in Maine correlates with Per capita consumption of margarine

Correlation: 99.26% ($r=0.992558$)



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

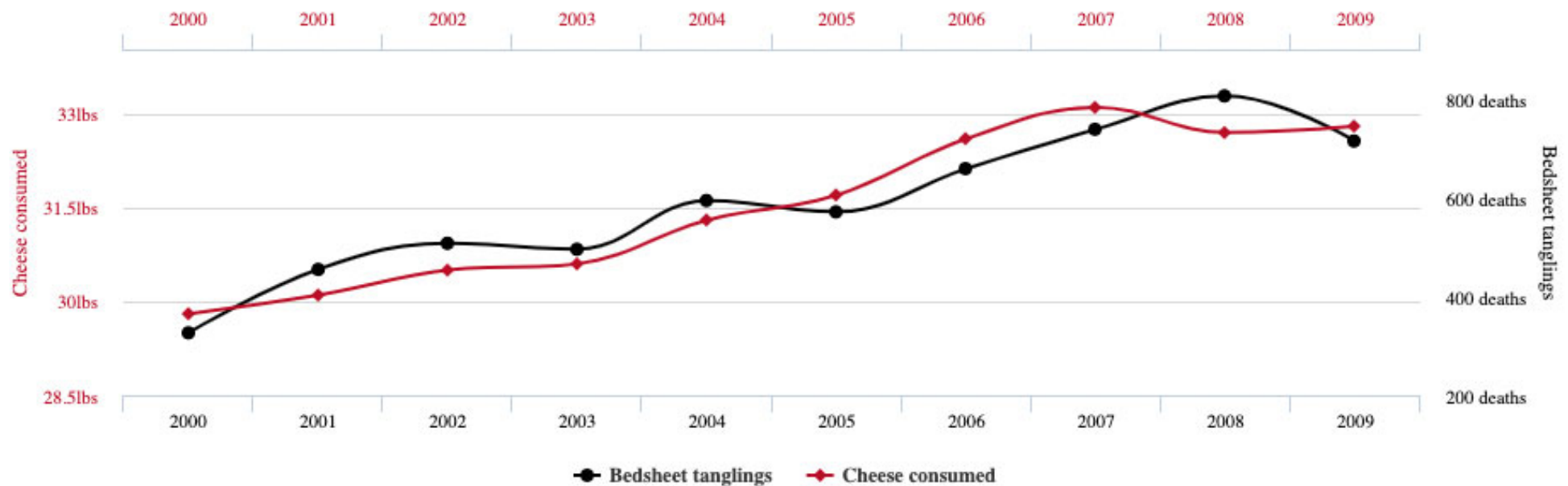
tylervigen.com

Per capita cheese consumption

correlates with

Number of people who died by becoming tangled in their bedsheets

Correlation: 94.71% ($r=0.947091$)

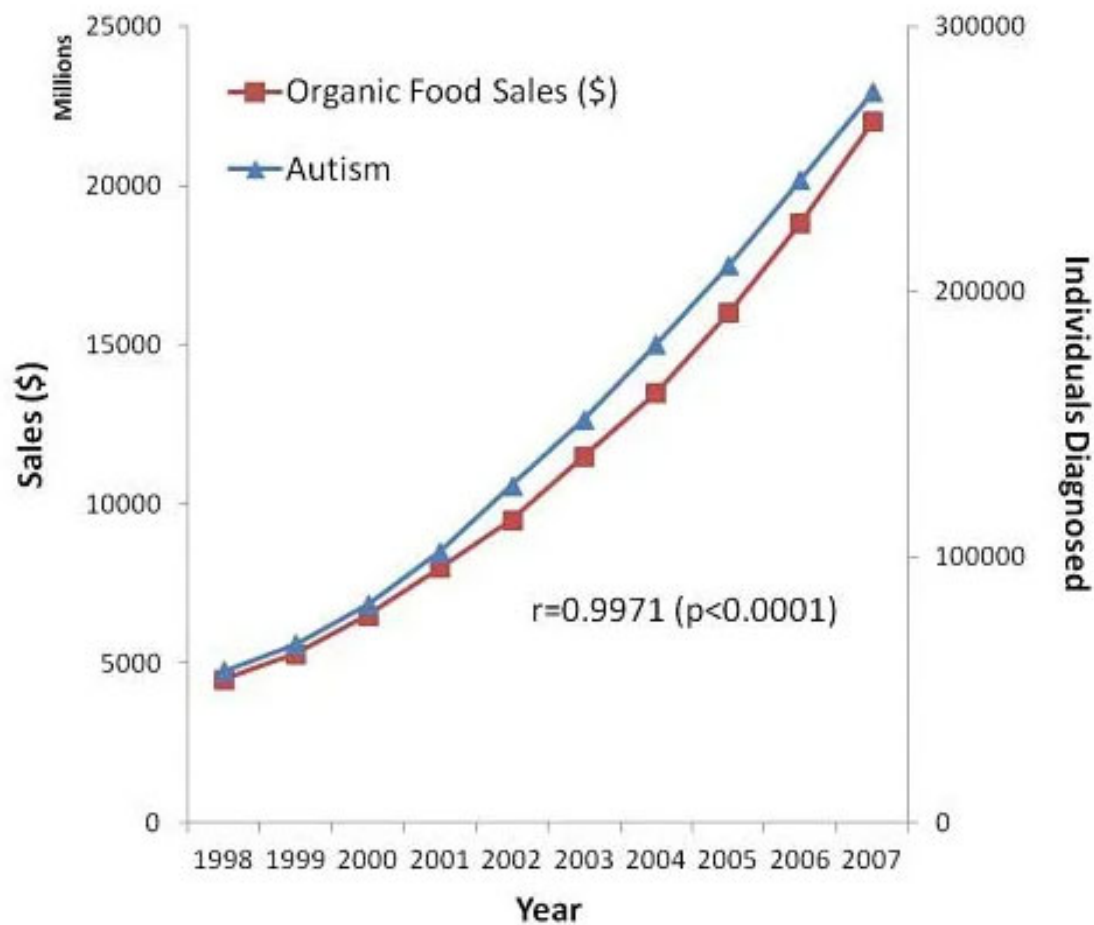


tylervigen.com

Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention

These examples come from someone who wrote a script to harvest data from a bunch of publicly accessible databases and plot a bunch of things against each other until something (accidentally) correlated strongly! ("p-hacking")

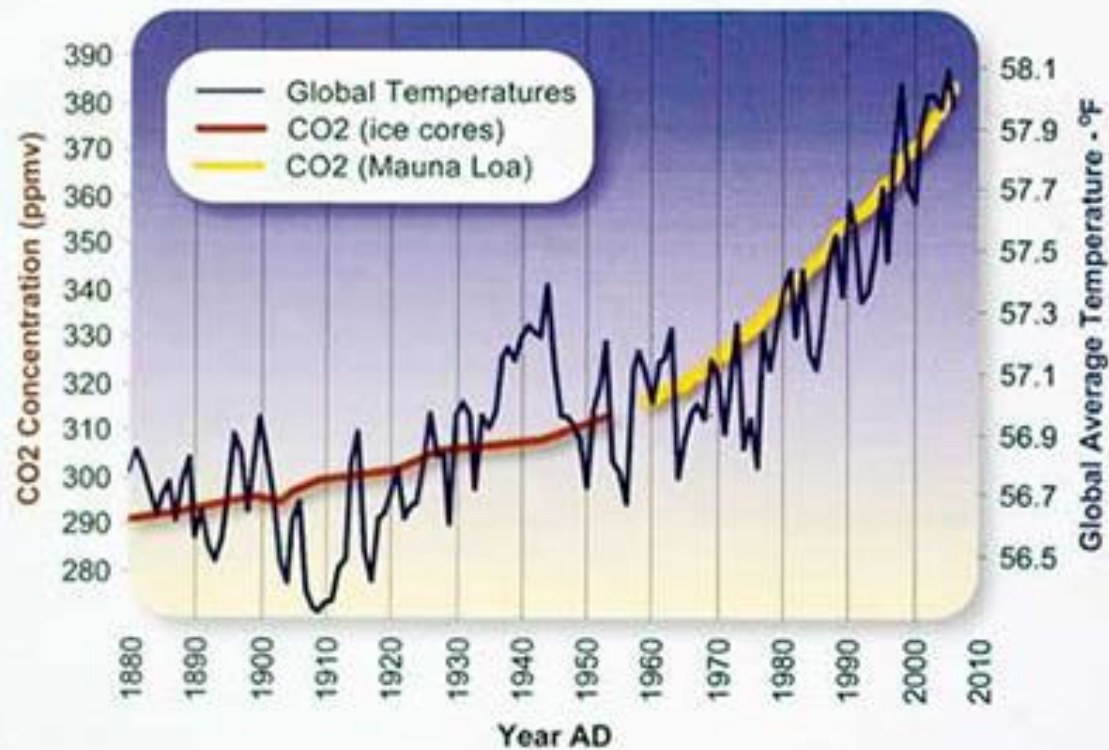
More fun correlations: <https://www.tylervigen.com/spurious-correlations>



This one is maybe not so funny – what do you do with correlations like this?

A correlation is the beginning of the study, not the end! If you see a correlation, can you design a follow-up study to test whether there is causation or not? Can you come up with any logical reasons for a causal relationship?

Global Average Temperature and Carbon Dioxide Concentrations, 1880-2006



Source: Michael E. Mann, Woods Hole Research Center, from *How We Know What We Know About Our Changing Climate*

This correlation, for example, has very robust theory logic confirming that it is causal. Also many controlled experiments on a much-smaller-than-planetary scale and some comparative planetary studies.