Government 10: Quantitative Political Analysis

Sean Westwood

Correlation

Correlation between variables

Correlation implies that two variables **occur** and/or **'move together'**:

Correlation between variables

Correlation implies that two variables **occur** and/or **'move together'**:

So, when we observe a change in X, we also observe a change in our outcome of interest: Y

Correlation Examples

$X \rightarrow Y$

Correlation Examples

$X \rightarrow Y$

"Undertaking two and a half hours a week of physical activity can reduce, but not banish, someone's chances of dying from any cause, according to the study of 36,370 British patients."

Correlation Examples

$X \rightarrow Y$

"Undertaking two and a half hours a week of physical activity can reduce, but not banish, someone's chances of dying from any cause, according to the study of 36,370 British patients."

Increased Exercise $\rightarrow$ Reduced Death Rate

Correlation Examples

$X \rightarrow Y$

"Pipe smoking...was associated with an increased risk of death from cancers of the lung.

Correlation Examples

$X \rightarrow Y$

"Pipe smoking...was associated with an increased risk of death from cancers of the lung. These risks were generally smaller than those associated with cigarette smoking and similar to or larger than those associated with cigar smoking"

Correlation Examples

$X \rightarrow Y$

"Pipe smoking...was associated with an increased risk of death from cancers of the lung. These risks were generally smaller than those associated with cigarette smoking and similar to or larger than those associated with cigar smoking"

Pipe smoking $\rightarrow$ Higher Risk of Death

Correlation Examples

$X \rightarrow Y$

"Pipe smoking...was associated with an increased risk of death from cancers of the lung. These risks were generally smaller than those associated with cigarette smoking and similar to or larger than those associated with cigar smoking"

Pipe smoking $\rightarrow$ Higher Risk of Death

Cigarette smoking $\rightarrow$ Lower Risk of Death

Correlation and Causation

Probe this logically...

Correlation and Causation

Probe this logically...

$$\text{Pipe Smoking} \rightarrow \text{Death}$$

Correlation and Causation

Probe this logically...

$$\text{Pipe Smoking} \rightarrow \text{Death}$$

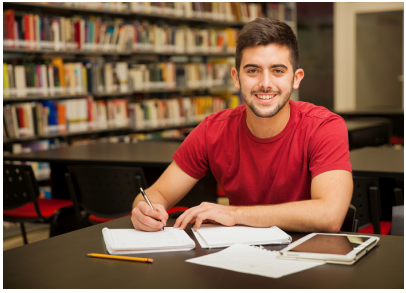
What could be wrong here?

Correlation and Causation

Who smokes a pipe?

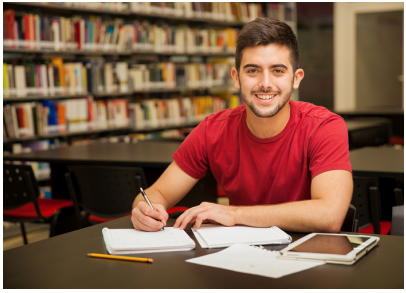
Correlation and Causation

Who smokes a pipe?



Correlation and Causation

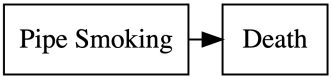
Who smokes a pipe?



or

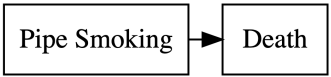
Correlation and Causation

Instead of:

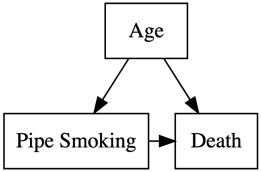


Correlation and Causation

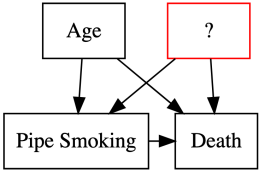
Instead of:



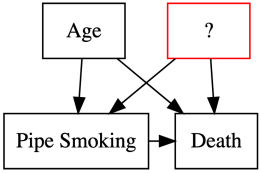
Perhaps:



Correlation and Causation

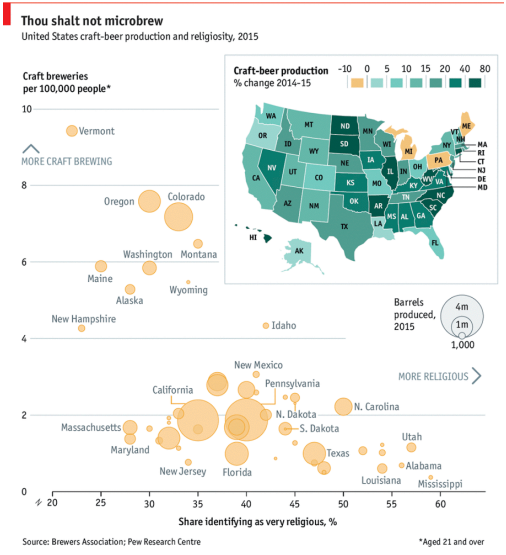


Correlation and Causation



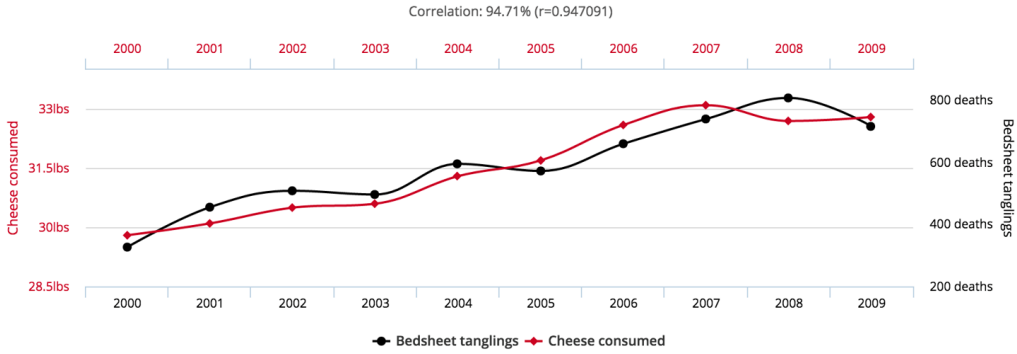
These (unmeasured) variables are known as **confounders**

Some examples of where relationships go wrong



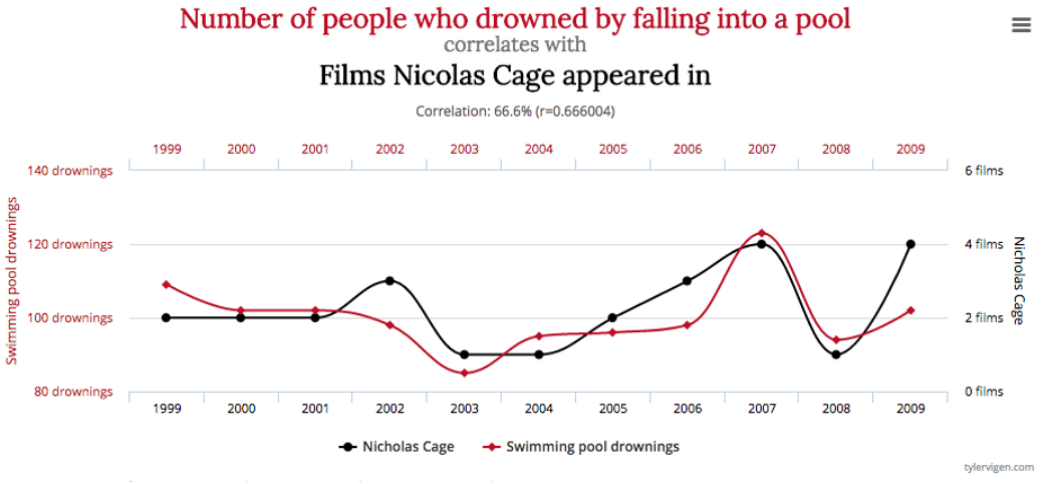
What else do these states have in common?

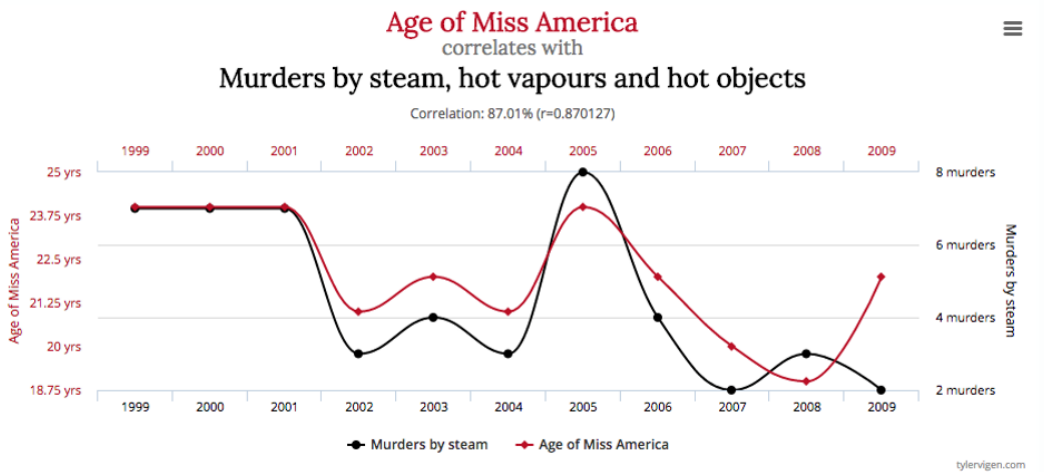
Per capita cheese consumption
correlates with
Number of people who died by becoming tangled in their bedsheets

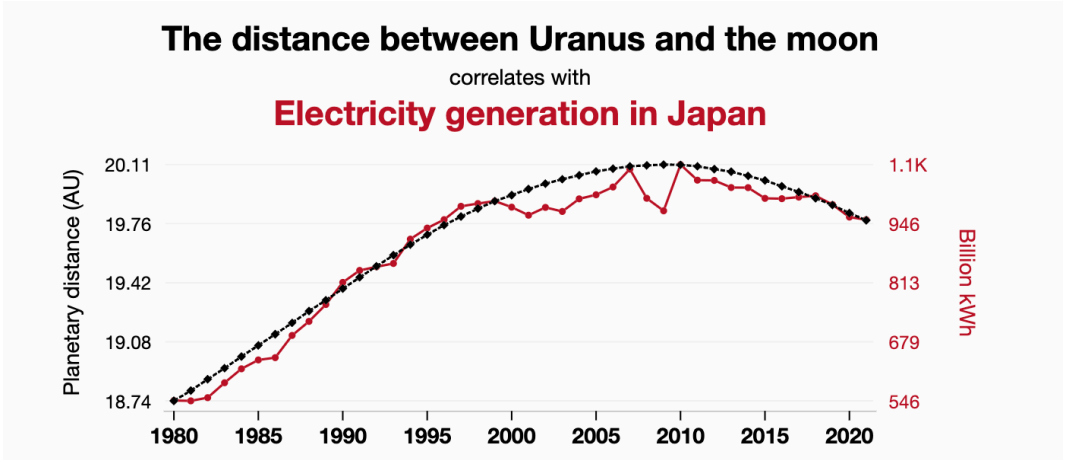


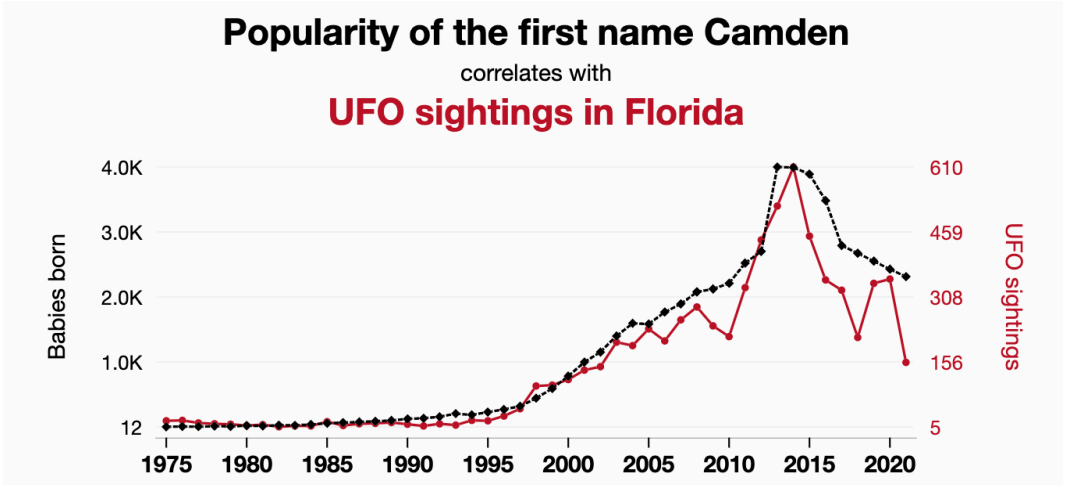
tylervigen.com

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


Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in
Correlation: 66.6% (r=0.666004)
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
140 drowningsSwimming pool drownings
120 drowningsSwimming pool drownings
100 drowningsSwimming pool drownings
80 drowningsSwimming pool drownings
6 filmsNicholas Cage
4 filmsNicholas Cage
2 filmsNicholas Cage
0 filmsNicholas Cage
Nicholas Cage
Swimming pool drownings
tylervigen.com


Age of Miss America
correlates with
Murders by steam, hot vapours and hot objects
Correlation: 87.01% (r=0.870127)
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
25 yrs
23.75 yrs
22.5 yrs
21.25 yrs
20 yrs
18.75 yrs
Age of Miss America
8 murders
6 murders
4 murders
2 murders
Murders by steam
Murders by steam
Age of Miss America
tylervigen.com


The distance between Uranus and the moon
correlates with
Electricity generation in Japan
Planetary distance (AU)
20.11
19.76
19.42
19.08
18.74
1.1K
946
813
679
546
Billion kWh
1980
1985
1990
1995
2000
2005
2010
2015
2020


Popularity of the first name Camden
correlates with
UFO sightings in Florida
Babies born
4.0K
3.0K
2.0K
1.0K
12
610
459
308
156
5
UFO sightings
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020

Correlation and Causation

Everything causal is correlated

Correlation and Causation

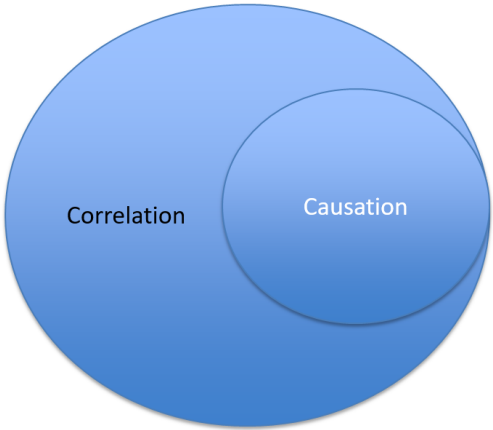
Everything causal is correlated

Not all correlation is causal

Correlation and Causation

Everything causal is correlated

Not all correlation is causal



Correlation and Causation Example

Research Question:
How much racial discrimination exists in the labor market?

Correlation and Causation Example

Research Question:
How much racial discrimination exists in the labor market?

Hypothesis:
African Americans are **less likely** to receive a response to job applications than white applicants.

Correlation and Causation Example

Research Question:
How much racial discrimination exists in the labor market?

Hypothesis:
African Americans are **less likely** to receive a response to job applications than white applicants.

How do we test this?

Correlation and Causation

There are two primary ways to address confounders:

Correlation and Causation

There are two primary ways to address confounders:

1. **Observational: 'Controlling'**

Correlation and Causation

There are two primary ways to address confounders:

1. **Observational: 'Controlling'**
   - For variables we can measure, implement designs so that the groups are comparable.

Correlation and Causation

There are two primary ways to address confounders:

1. **Observational: 'Controlling'**
   - For variables we can measure, implement designs so that the groups are comparable.
2. **Experiment: Randomization**

Correlation and Causation

There are two primary ways to address confounders:

1. **Observational: 'Controlling'**
   - For variables we can measure, implement designs so that the groups are comparable.
2. **Experiment: Randomization**
   - With a sufficient sample size, random assignment assures that groups are comparable on unobserved factors.

Research Design A - Observational

1. Solicit 1,000 volunteers that identify as white, and 1,000 volunteers that identify as black.

Research Design A - Observational

1. Solicit 1,000 volunteers that identify as white, and 1,000 volunteers that identify as black.
2. Ask these volunteers to send their resumes to a single job opening, and report back whether they received a response.

Research Design A - Observational

1. Solicit 1,000 volunteers that identify as white, and 1,000 volunteers that identify as black.
2. Ask these volunteers to send their resumes to a single job opening, and report back whether they received a response.

Research Design B - Experimental

1. Create identical resumes. Randomly assign half of the resumes to contain a 'white' cue, and half to contain a 'black' cue (name).

Research Design B - Experimental

1. Create identical resumes. Randomly assign half of the resumes to contain a 'white' cue, and half to contain a 'black' cue (name).
2. Randomly sample thousands of employers.

Research Design B - Experimental

1. Create identical resumes. Randomly assign half of the resumes to contain a 'white' cue, and half to contain a 'black' cue (name).
2. Randomly sample thousands of employers.
3. Send exactly one resume to each employer.
4. Measure responses

Differences?

Which gives the most precise answer?

Which is the easiest?