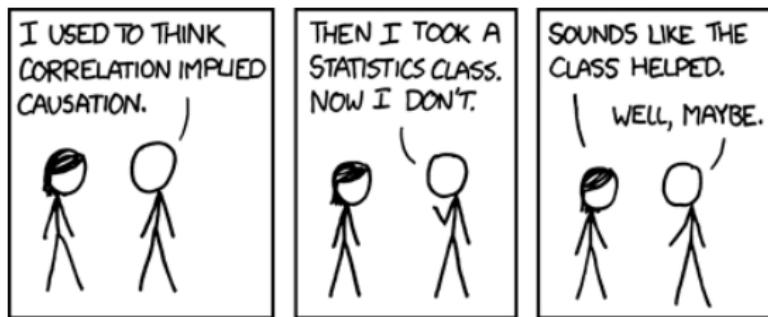


CORRELATION VS. CAUSATION

Data Analysis for Journalism and Political Communication
(Fall 2025)

Prof. Bell



"CONVICTION" BY RACHEL AVIV (NEW YORKER)

Operation Hummingbird

Exhibit Ref: CEH/16

Staff Presence Report

Chart 1: Nurses present on clinical and administrative duties Summary

Event	Baby	Date Shift Started	Shift Type	Asha SIMPSON	Nursing Staff on Clinical and Administrative Duties																																				
					Angela MISHANE	Ashleigh HUDDON	Belinda SIMCOCK	Bernadette BUTTERWORTH	Caroline BENNIION	Caroline DAKLEY	Cheryl CLIFFBERTSON-TAYLOR	Christopher BOOTH	Erian POWELL	Elizabeth MARSHALL	Janet COX	Jean PIERES	Jennifer JONES-KAY	Joanne WILLIAMS	Kathryn WARD	Laura EAGLES	Lisa WALTER	Lucy LEFFERY	Mary GRIFFITH	Melanie TAYLOR	Menna LAPALAMEN	Nicola DENNISON	Patricia STEELE	Rebecca MORIGAN	Samantha O'BRIEN	Shelley TOMLIN	Sophie ELLIS	Valerie PARKES	Valerie THOMAS	Vicky BLAINE	Yvonne FARMER	Yvonne GRIFFITHS					
1	Child A	08/06/2015	NIGHT						X				X								X	X	X	X																	
2	Child B	09/06/2015	NIGHT							X											X	X	X																		
3	Child C	13/06/2015	NIGHT								X										X																				
4	Child D	21/06/2015	NIGHT								X	X		X							X																				
5	Child E	03/08/2015	NIGHT								X		X									X	X																		
6	Child F	04/08/2015	NIGHT								X			X								X																			
7	Child G	06/09/2015	NIGHT	X																	X	X																			
8	Child G	21/09/2015	DAY								X	X		X	X	X					X																				
9	Child H	25/09/2015	NIGHT																			X																			
10	Child H	26/09/2015	NIGHT																			X																			
11	Child I	30/09/2015	DAY								X	X		X	X	X					X	X	X	X	X																
12	Child I	12/10/2015	NIGHT									X										X	X																		
13	Child I	13/10/2015	NIGHT									X			X	X	X					X																			
14	Child I	22/10/2015	NIGHT									X										X																			
15	Child J	26/11/2015	NIGHT									X										X	X	X	X																
16	Child J	17/12/2015	DAY									X			X	X	X				X	X	X	X	X																
17	Child K	16/02/2016	NIGHT									X										X																			
18	Child M	09/04/2016	DAY									X	X	X								X	X																		
19	Child L	09/04/2016	DAY									X										X	X																		
20	Child N	02/06/2016	NIGHT										X									X	X																		
21	Child N	14/06/2016	NIGHT										X	X							X	X	X	X																	
22	Child O	23/06/2016	DAY										X		X							X		X	X	X	X														
23	Child P	23/06/2016	DAY									X		X		X						X	X	X	X	X															
24	Child P	24/06/2016	DAY									X			X							X	X		X																
25	Child Q	25/06/2016	DAY											X								X	X	X	X																
		Total	1	1	1	2	5	5	2	2	3	7	2	7	3	5	7	6	2	3	3	3	2	4	6	25	7	7	2	3	6	3	2	3	4	4	5	6	2	6	5

*('X' indicates 'on duty' presence on the shift, where a suspicious event has been identified)

Correlation

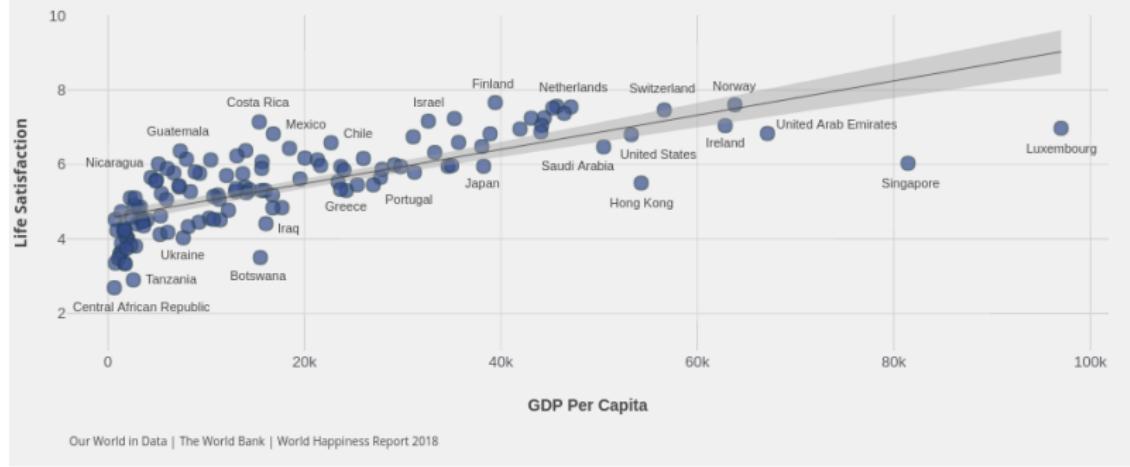
Correlation exists when the absolute rate of change in the values of two variables are similar.

Correlation

Correlation exists when the absolute rate of change in the values of two variables are similar.

- **Positive correlation:** As the value of one variable increases (decreases), the value of the other variable increases (decreases) at the same rate

Residents of countries with higher **national incomes** tend to report higher **life satisfaction**.



Correlation

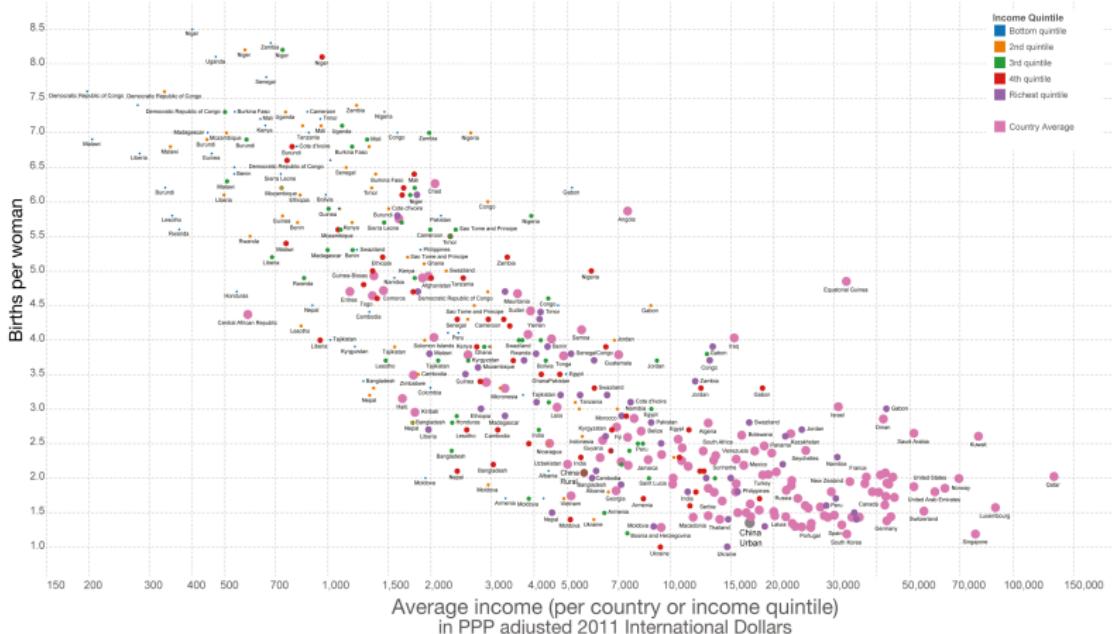
Correlation exists when the absolute rate of change in the values of two variables are similar.

- **Positive correlation:** As the value of one variable increases (decreases), the value of the other variable increases (decreases) at the same rate
- **Negative correlation:** As the value of one variable increases (decreases), the value of the other variable decreases (increases) at the same rate

Births per woman by income level, 2013

Pink bubbles show country averages for income (GDP per capita, PPP adjusted) and for the total fertility rate.

For all other countries the fertility rate is shown for each wealth quintile within the country. It is plotted against the average income per corresponding quintile in the same country.



Data sources: World Bank for all income measures. Fertility rates: national averages from WDI. Fertility by wealth quintile from the DHS (via the WHO) – except for China for which data was added from various research papers. Most data are from 2013 – none of the data refer to a year earlier than 2005.

Licensed under CC-BY-SA by the author Max Roser.

Causation

Causation exists when a change in the value of one variable would not be observed without a preceding change in the value of another variable.

Causation

Causation exists when a change in the value of one variable would not be observed without a preceding change in the value of another variable.

- Correlation is descriptive, while causation is predictive

Causation

Causation exists when a change in the value of one variable would not be observed without a preceding change in the value of another variable.

- Correlation is descriptive, while causation is predictive
- We say that the **independent** or **explanatory** variable causes the **dependent** or **outcome** variable

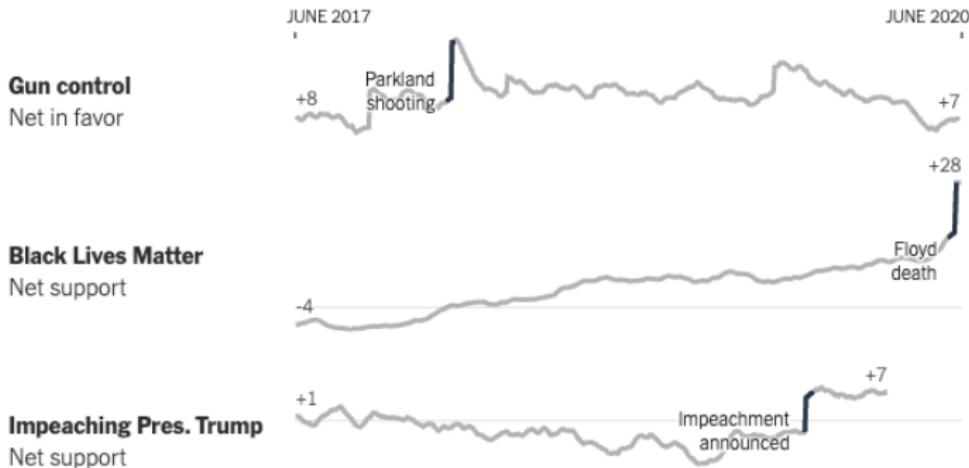
Causation

Causation exists when a change in the value of one variable would not be observed without a preceding change in the value of another variable.

- Correlation is descriptive, while causation is predictive
- We say that the **independent** or **explanatory** variable causes the **dependent** or **outcome** variable
- Causation depends on knowing the **counterfactual**: if we did not observe a change in the value of the explanatory variable, we would not observe a change in the value of the outcome variable

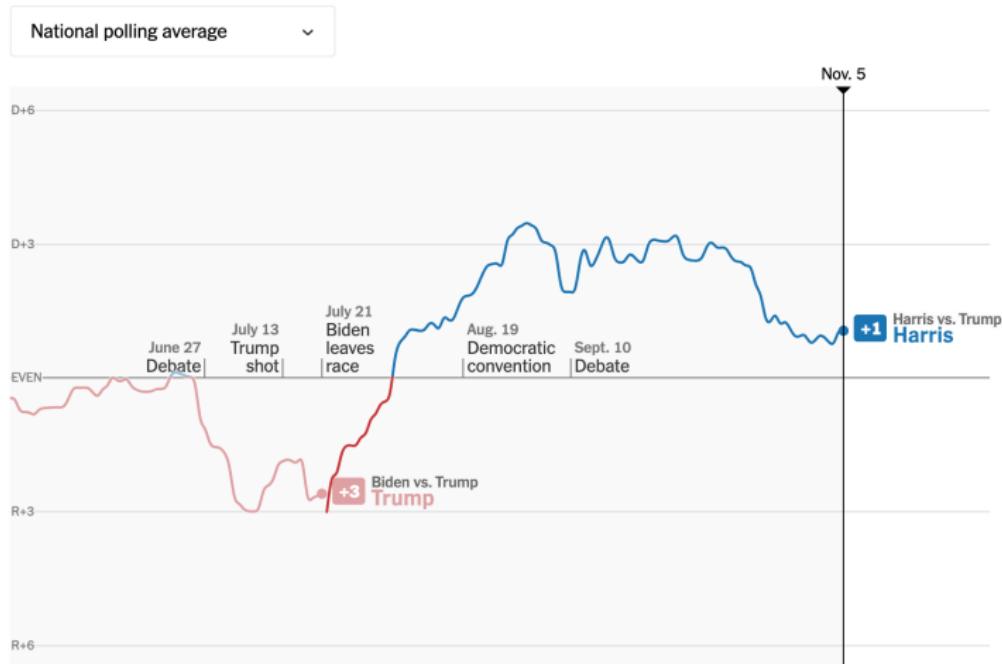
How voters' views on other issues have changed in the last two years

Large swings in public opinion in short periods are not typical. Two-week periods with the biggest shifts in movement are highlighted.



From Biden to Harris

This chart shows how the polling margin has changed over the course of the campaign, first for the Biden vs. Trump matchup, and now for Harris vs. Trump.



Note: Head-to-head average shown for the Biden vs. Trump matchup. The Harris vs. Trump average includes polls conducted before Biden dropped out and polls that included Robert F. Kennedy Jr.

“15 Lessons Scientists Learned About Us When the World Stood Still” (New York Times)

- ① Home-field advantage is real (when fans are there).
- ② Women who rest more and were exposed to fewer stressors, pollutants and viruses have fewer premature births.
- ③ Dolphins whistle longer, birds change their songs, and sea turtles lay more eggs when humans aren't around.

The Fundamental Problem of Causal Inference

For any given case, we observe the outcome variable with either a change in the independent variable or no change in the independent variable, but not both.

The Fundamental Problem of Causal Inference

For any given case, we observe the outcome variable with either a change in the independent variable or no change in the independent variable, but not both.

- In other words, we do not observe the counterfactual

The Fundamental Problem of Causal Inference

For any given case, we observe the outcome variable with either a change in the independent variable or no change in the independent variable, but not both.

- In other words, we do not observe the counterfactual
- But we try our best to observe the counterfactual using **experiments**

The Fundamental Problem of Causal Inference

For any given case, we observe the outcome variable with either a change in the independent variable or no change in the independent variable, but not both.

- In other words, we do not observe the counterfactual
- But we try our best to observe the counterfactual using **experiments**
- Experiments establish a counterfactual by comparing cases that differ only in the explanatory variable that we are interested in

Control



Experiment



Identical pots



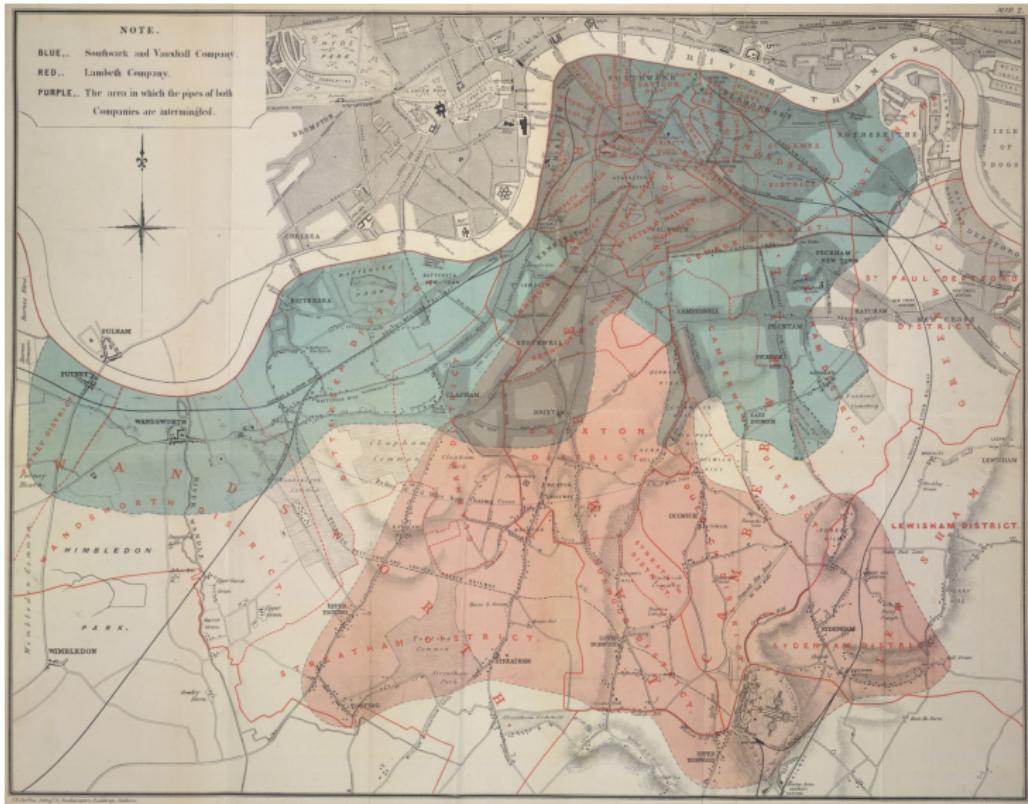
Fertilizer is
independent variable



Plant growth is
dependent variable



JOHN SNOW'S CHOLERA EXPERIMENT



JOHN SNOW'S CHOLERA EXPERIMENT

	Number of houses.	Deaths from Cholera.	Deaths in each 10,000 houses.
Southwark and Vauxhall Company	40,046	1,263	315
Lambeth Company . . .	26,107	98	37
Rest of London . . .	256,423	1,422	59

EXPERIMENTS

- A critical element of experiments is that the cases are assigned to the **treatment** group (e.g., getting a drug) and **control** group (e.g., getting a placebo) completely at random

EXPERIMENTS

- A critical element of experiments is that the cases are assigned to the **treatment** group (e.g., getting a drug) and **control** group (e.g., getting a placebo) completely at random
- Recall the definition of a **random sample**: the probability of any given unit being drawn from the population is uniform (the same)

EXPERIMENTS

- A critical element of experiments is that the cases are assigned to the **treatment** group (e.g., getting a drug) and **control** group (e.g., getting a placebo) completely at random
- Recall the definition of a **random sample**: the probability of any given unit being drawn from the population is uniform (the same)
- The intuition for randomness is that there is no **selection bias**: patients aren't getting the drug because they are younger or healthier, for example.

EXPERIMENTS

- But wait: were households in London assigned to either the Southwark & Vauxhall Company or the Lambeth Company completely at random?

EXPERIMENTS

- But wait: were households in London assigned to either the Southwark & Vauxhall Company or the Lambeth Company completely at random?
- No. This is called a natural experiment, and it relies on whether we believe being in either group is as-good-as-random.

EXPERIMENTS

Card and Krueger (1994)

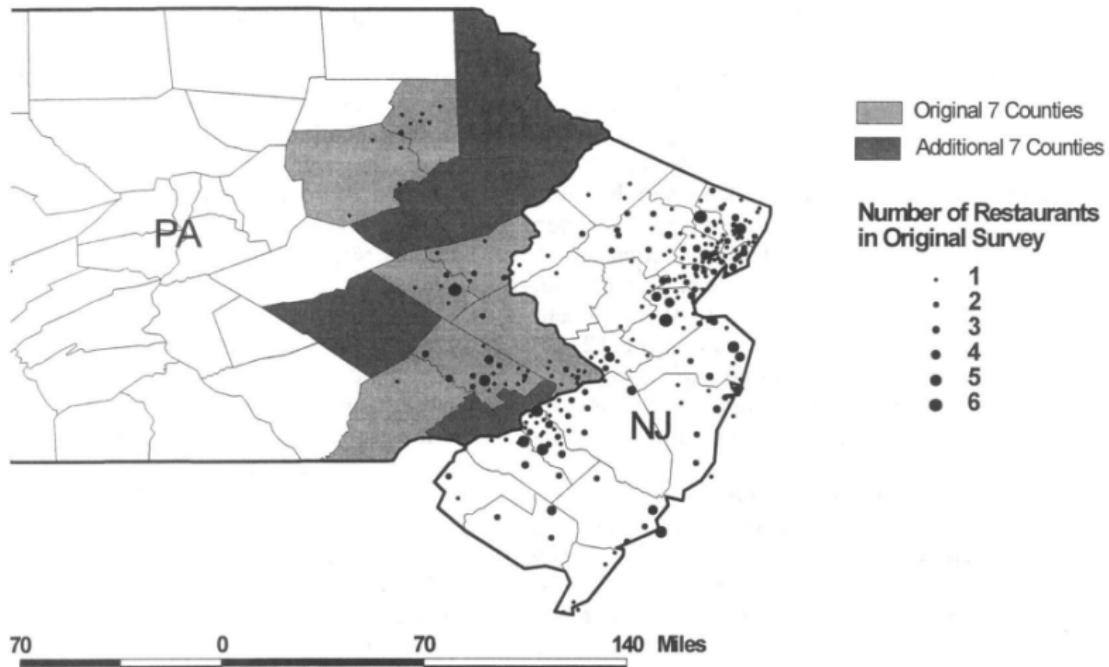


FIGURE 1. AREAS OF NEW JERSEY AND PENNSYLVANIA COVERED BY ORIGINAL SURVEY AND BLS DATA

EXPERIMENTS

- But wait: were households in London assigned to either the Southwark & Vauxhall Company or the Lambeth Company completely at random?
- No. This is called a natural experiment, and it relies on whether we believe being in either group is as-good-as-random.
- Finding natural experiments in the real world is really difficult, so we often design our experiments in controlled settings like laboratories or surveys

EXPERIMENTS

- But wait: were households in London assigned to either the Southwark & Vauxhall Company or the Lambeth Company completely at random?
- No. This is called a natural experiment, and it relies on whether we believe being in either group is as-good-as-random.
- Finding natural experiments in the real world is really difficult, so we often design our experiments in controlled settings like laboratories or surveys
- We only have a true experiment where the researcher randomly assigns cases to treatment or control.

EXPERIMENTS

Kam and Zechmeister (2013)



The logo consists of two thick horizontal blue bars. Between them, the name "BEN GRIFFIN" is displayed in large, bold, sans-serif capital letters. "BEN" is in red, and "GRIFFIN" is in blue.

ASSESSING CAUSALITY WITHOUT EXPERIMENTS

- What about when we don't have an experiment, but instead are collecting **observational data** from the world?

ASSESSING CAUSALITY WITHOUT EXPERIMENTS

- What about when we don't have an experiment, but instead are collecting **observational data** from the world?
- Then we face the fundamental problem of causal inference. Correlation does not imply causation.

ASSESSING CAUSALITY WITHOUT EXPERIMENTS

- What about when we don't have an experiment, but instead are collecting **observational data** from the world?
- Then we face the fundamental problem of causal inference. Correlation does not imply causation.
- But all hope is not lost - we just have to be much more careful about threats to causal inference: **confounders** and **reverse causation**.

CONFOUNDERS

Confounder

A variable that explains change in both the independent and dependent variable.

CONFOUNDERS

Confounder

A variable that explains change in both the independent and dependent variable.

Views of the
Economy



Presidential
Vote Intention

CONFOUNDERS

Confounder

A variable that explains change in both the independent and dependent variable.

Views of the
Economy

Presidential
Vote Intention



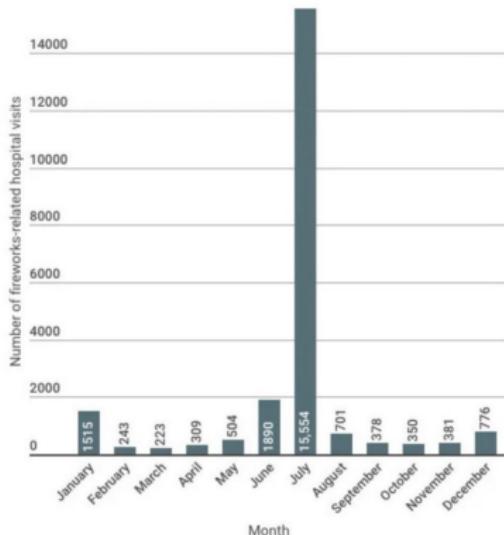
Media
Coverage of
Economy

CONFOUNDERS

"It's most dangerous to launch fireworks in July."

IV: Time

DV: Fireworks-related hospitalizations

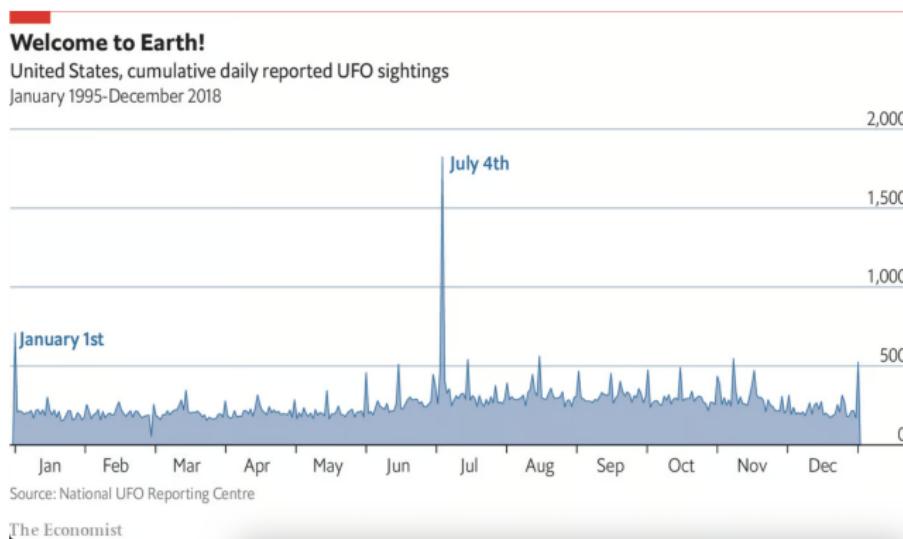


CONFOUNDERS

"If aliens exist, they will arrive on July 4th."

IV: Time

DV: UFO Sightings



REVERSE CAUSATION

Reverse Causation

Occurs when the dependent variable affects the independent variable.

REVERSE CAUSATION

Reverse Causation

Occurs when the dependent variable affects the independent variable.

Views of the
Economy



Presidential
Vote Intention

REVERSE CAUSATION

Reverse Causation

Occurs when the dependent variable affects the independent variable.

Views of the
Economy

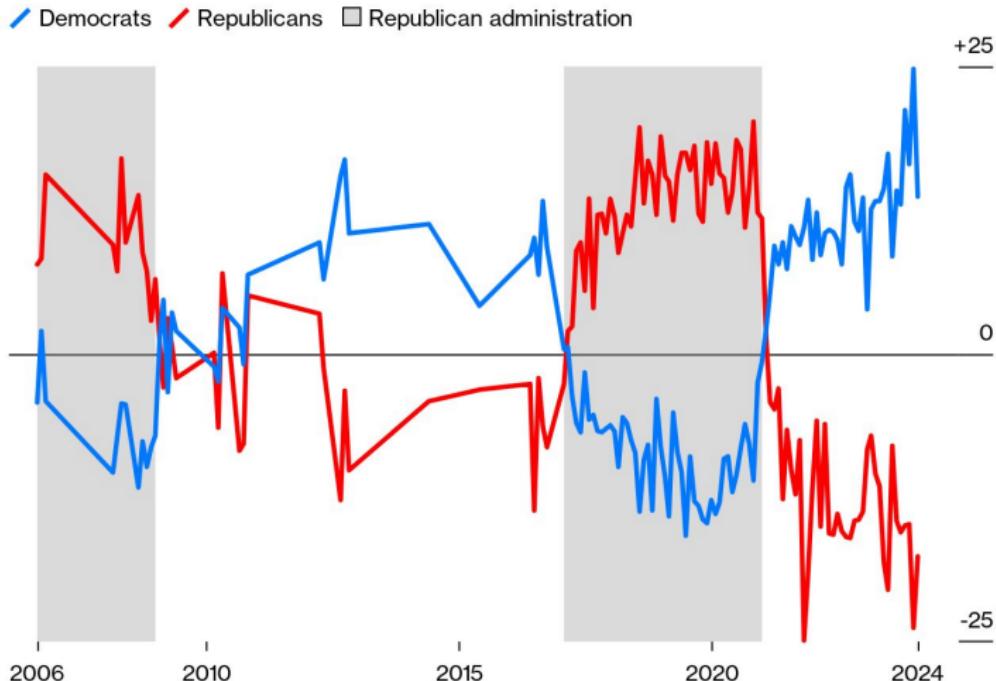


Presidential
Vote Intention

REVERSE CAUSATION

Current Consumer Sentiment

Difference between current consumer sentiment among all respondents and respondents by political party



Source: University of Michigan Index of Consumer Sentiment

Bloomberg Businessweek

EXERCISE: CORRELATION VS. CAUSATION