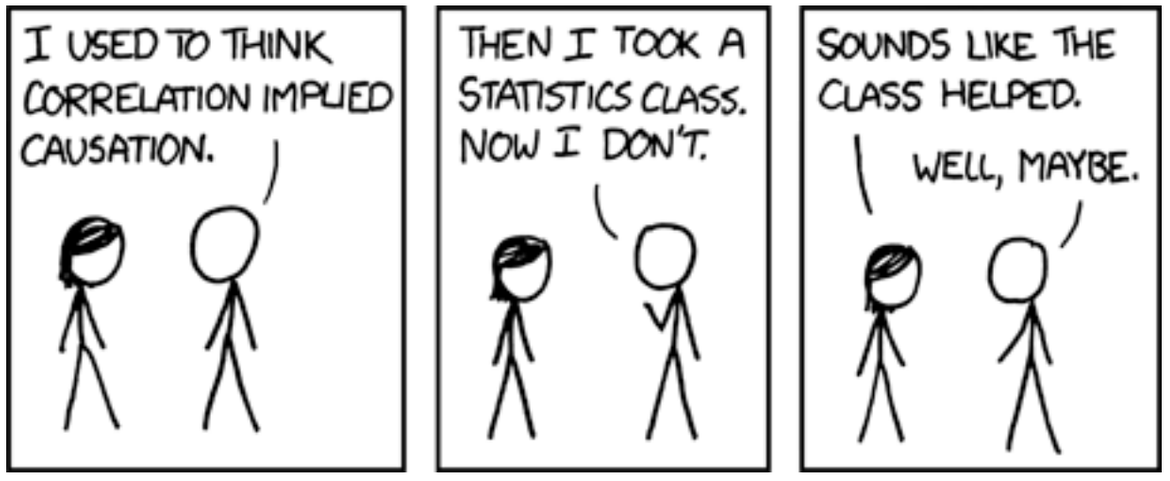
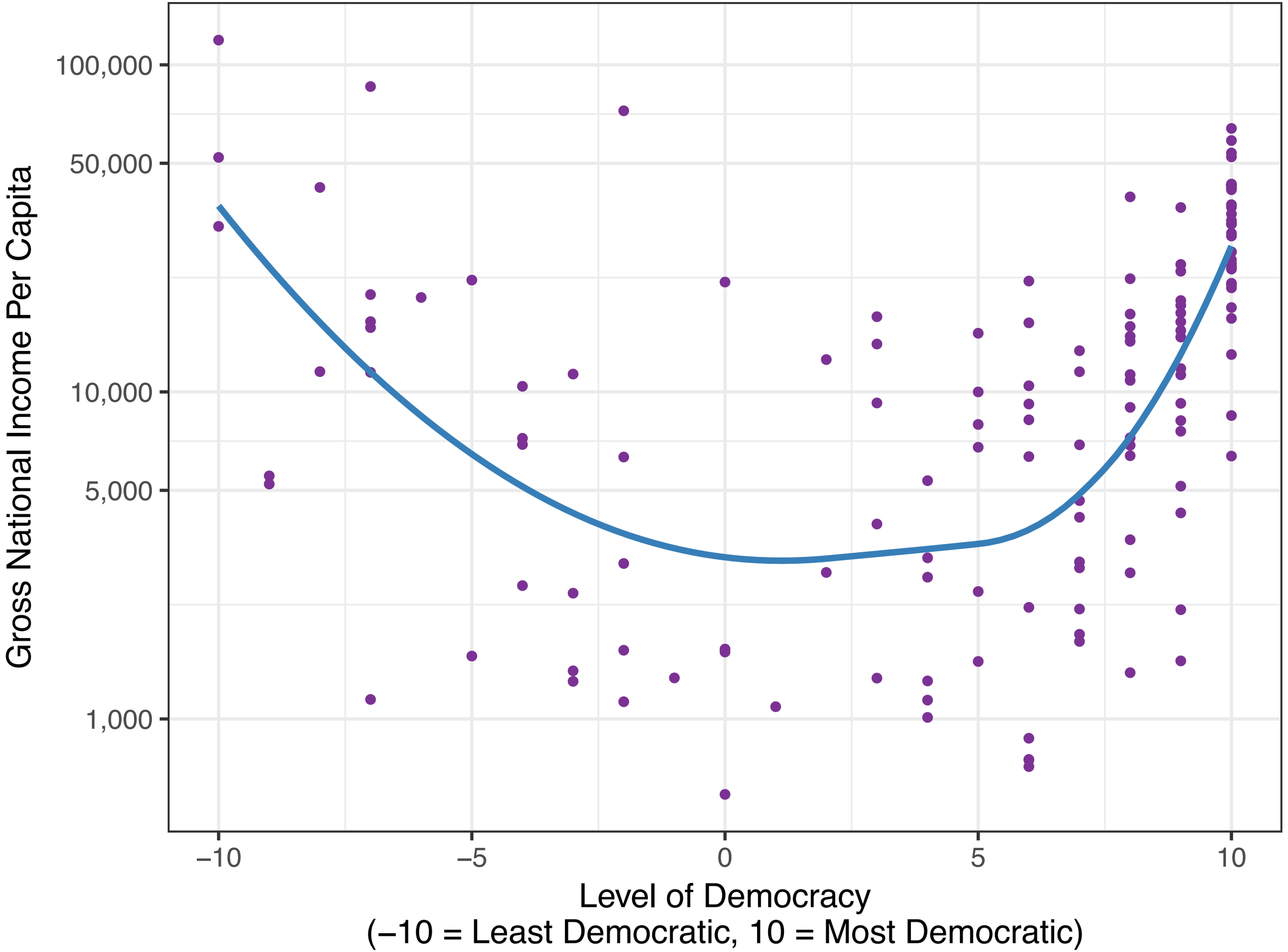
A Partial Solution

To the Fundamental Problem of Causal Inference



Some of our most important questions are **causal questions**.


Gross National Income Per Capita
100,000
50,000
10,000
5,000
1,000
−10
−5
0
5
10
Level of Democracy
(−10 = Least Democratic, 10 = Most Democratic)

Gross National Income Per Capita

100,000
50,000
10,000
5,000
1,000

Qatar
Kuwait
Singapore
Saudi Arabia
Oman
Bahrain
Equatorial Guinea
Libya
Cuba
Belarus
Kazakhstan
Azerbaijan
Turkmenistan
Jordan
Egypt
Fiji
Morocco
Angola
Swaziland
Uzbekistan
Norway
Luxembourg
United States
Switzerland
Austria
Germany
Denmark
Netherlands
Canada
Australia
Belgium
United Kingdom
France
Italy
Sweden
Czech Republic
New Zealand
Finland
Ireland
Malaysia
Trinidad and Tobago
Israel
Slovenia
Spain
Latvia
Croatia
Estonia
Cyprus
Argentina
Portugal
Mexico
Poland
Turkey
Gabon
Botswana
Greece
Lebanon
Brazil
Romania
Hungary
Iraq
Suriname
Thailand
Lithuania
Montenegro
Panama
Bulgaria
Chile
Algeria
Ecuador
Tunisia
Colombia
Serbia
Peru
South Africa
Sri Lanka
Namibia
Dominican Republic
Uruguay
Armenia
El Salvador
Indonesia
Mauritius
Ukraine
Georgia
Albania
Bhutan
Guyana
Guatemala
Costa Rica
Nigeria
Jamaica
Pakistan
Paraguay
Mongolia
Yemen
Djibouti
Kyrgyzstan
Cape Verde
Honduras
India
Mauritania
Cambodia
Papua New Guinea
Cameroon
Zambia
Philippines
Nicaragua
Tajikistan
Bangladesh
Nepal
Ghana
Lesotho
Gambia
Haiti
Chad
Mali
Senegal
Rwanda
Burkina Faso
Madagascar
Zimbabwe
Sierra Leone
Kenya
Eritrea
Ethiopia
Uganda
Guinea–Bissau
Guinea
Benin
Solomon Islands
Comoros
Togo
Mozambique
Niger
Burundi
Liberia
Central African Republic
Malawi

–10
–5
0
5
10

Level of Democracy
(–10 = Least Democratic, 10 = Most Democratic)

Gross National Income Per Capita

100,000
50,000
10,000
5,000
1,000

Qatar
Kuwait
Singapore
Saudi Arabia
Oman
Bahrain
Equatorial Guinea
Libya
Cuba
Belarus
Kazakhstan
Azerbaijan
Turkmenistan
China
Jordan
Egypt
Fiji
Morocco
Angola
Swaziland
Uzbekistan
Mauritania
Cameroon
Tajikistan
Gambia
Chad
Haiti
Rwanda
Burkina Faso
Eritrea
Ethiopia
Uganda
Togo
Guinea–Bissau
Central African Republic
Madagascar
Zimbabwe
Guinea
Mozambique
Niger
Burundi
Liberia
Malawi
Norway
Luxembourg
United States
Switzerland
Austria
Germany
Denmark
Netherlands
Canada
Australia
Belgium
United Kingdom
France
Sweden
Italy
Czech Republic
New Zealand
Finland
Ireland
Malaysia
Trinidad and Tobago
Israel
Slovenia
Spain
Latvia
Croatia
Estonia
Cyprus
Argentina
Portugal
Mexico
Poland
Turkey
Gabon
Botswana
Greece
Lebanon
Brazil
Romania
Hungary
Iraq
Suriname
Thailand
Lithuania
Montenegro
Panama
Bulgaria
Chile
Algeria
Ecuador
Tunisia
Colombia
Serbia
Peru
South Africa
Sri Lanka
Namibia
Dominican Republic
Uruguay
Armenia
El Salvador
Indonesia
Mauritius
Ukraine
Georgia
Albania
Bhutan
Guyana
Guatemala
Costa Rica
Nigeria
Jamaica
Pakistan
Paraguay
Mongolia
Yemen
Djibouti
Kyrgyzstan
Cape Verde
Honduras
India
Papua New Guinea
Zambia
Philippines
Nicaragua
Bangladesh
Nepal
Ghana
Lesotho
Mali
Senegal
Sierra Leone
Kenya
Benin
Solomon Islands
Comoros

−10
−5
0
5
10

Level of Democracy
(−10 = Least Democratic, 10 = Most Democratic)

Gross National Income Per Capita

100,000
50,000
10,000
5,000
1,000

Qatar
Kuwait
Saudi Arabia
Oman
Bahrain
Singapore
Libya
Equatorial Guinea
Cuba
Belarus
Kazakhstan
Azerbaijan
Turkmenistan
China
Jordan
Egypt
Fiji
Morocco
Angola
Swaziland
Uzbekistan
Norway
Luxembourg
United States
Switzerland
Austria
Germany
Denmark
Netherlands
Canada
Australia
Belgium
United Kingdom
France
Italy
Sweden
Czech Republic
New Zealand
Finland
Ireland
Malaysia
Trinidad and Tobago
Israel
Slovenia
Spain
Latvia
Croatia
Estonia
Cyprus
Argentina
Portugal
Mexico
Poland
Turkey
Gabon
Botswana
Greece
Lebanon
Brazil
Romania
Hungary
Iraq
Suriname
Thailand
Lithuania
Montenegro
Panama
Bulgaria
Chile
Algeria
Ecuador
Tunisia
Colombia
Serbia
Peru
South Africa
Sri Lanka
Namibia
Dominican Republic
Uruguay
Armenia
El Salvador
Indonesia
Mauritius
Ukraine
Georgia
Albania
Bhutan
Guyana
Guatemala
Costa Rica
Nigeria
Jamaica
Mongolia
Pakistan
Paraguay
Yemen
Djibouti
Kyrgyzstan
Cape Verde
Honduras
India
Mauritania
Cambodia
Papua New Guinea
Cameroon
Zambia
Philippines
Nicaragua
Tajikistan
Bangladesh
Nepal
Ghana
Lesotho
Gambia
Haiti
Chad
Mali
Senegal
Rwanda
Burkina Faso
Madagascar
Zimbabwe
Sierra Leone
Kenya
Eritrea
Ethiopia
Uganda
Guinea–Bissau
Guinea
Benin
Solomon Islands
Comoros
Togo
Mozambique
Niger
Burundi
Liberia
Central African Republic
Malawi

–10
–5
0
5
10

Level of Democracy
(–10 = Least Democratic, 10 = Most Democratic)

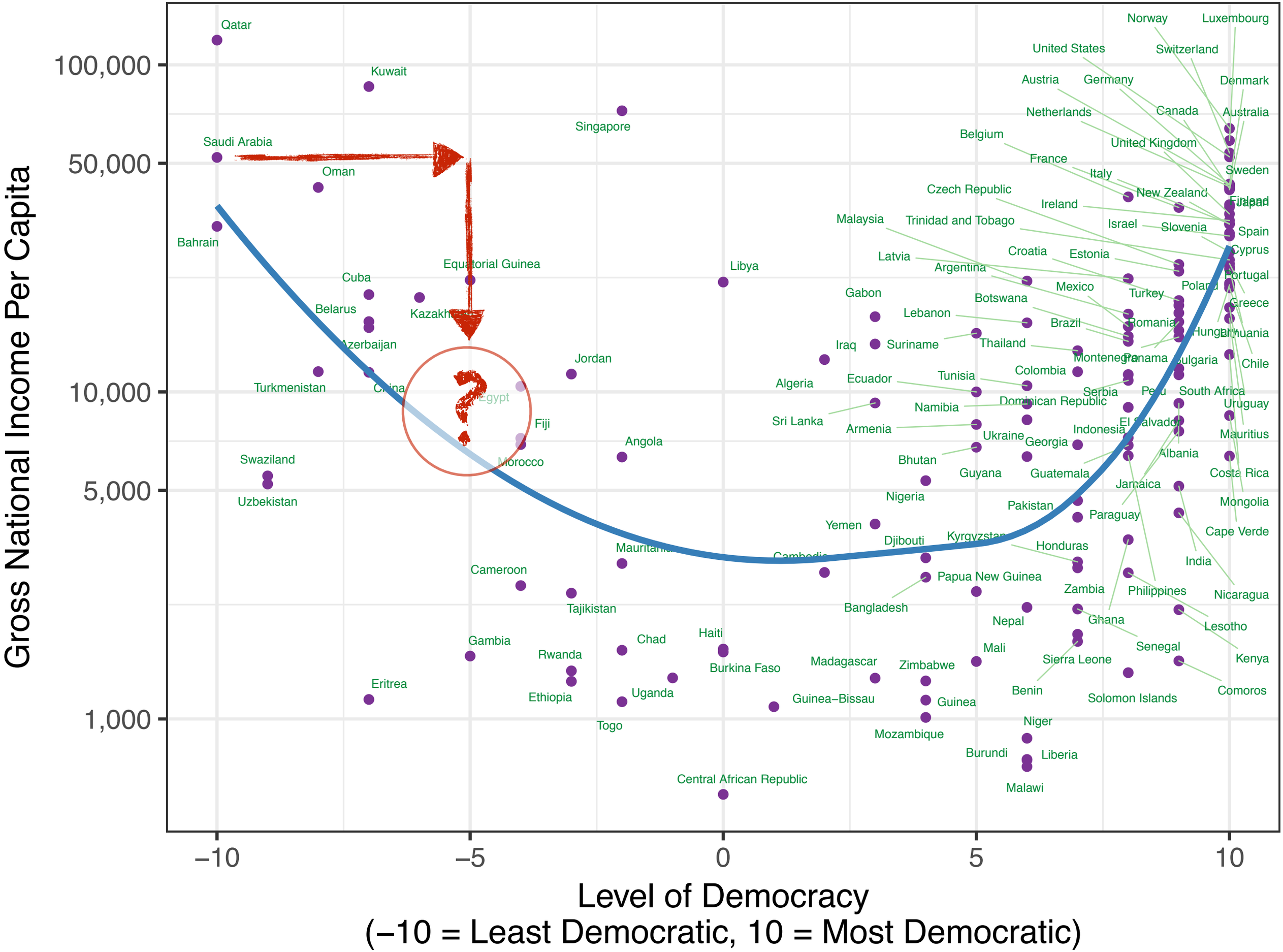
Gross National Income Per Capita

100,000
50,000
10,000
5,000
1,000

Qatar
Kuwait
Singapore
Saudi Arabia
Oman
Bahrain
Equatorial Guinea
Libya
Cuba
Belarus
Kazakhstan
Azerbaijan
Turkmenistan
China
Jordan
Egypt
Fiji
Morocco
Angola
Swaziland
Uzbekistan
Mauritania
Cameroon
Tajikistan
Gambia
Chad
Haiti
Rwanda
Burkina Faso
Eritrea
Ethiopia
Uganda
Togo
Guinea–Bissau
Central African Republic
Norway
Luxembourg
United States
Switzerland
Austria
Germany
Denmark
Netherlands
Canada
Australia
Belgium
United Kingdom
France
Italy
Sweden
Czech Republic
New Zealand
Finland
Ireland
Malaysia
Trinidad and Tobago
Israel
Slovenia
Spain
Latvia
Croatia
Estonia
Cyprus
Argentina
Portugal
Mexico
Poland
Gabon
Turkey
Botswana
Greece
Lebanon
Brazil
Romania
Hungary
Iraq
Suriname
Thailand
Lithuania
Montenegro
Panama
Bulgaria
Chile
Algeria
Ecuador
Tunisia
Colombia
Serbia
Peru
South Africa
Sri Lanka
Namibia
Dominican Republic
Uruguay
Armenia
El Salvador
Indonesia
Mauritius
Ukraine
Georgia
Albania
Bhutan
Guyana
Guatemala
Costa Rica
Nigeria
Jamaica
Mongolia
Pakistan
Paraguay
Yemen
Djibouti
Kyrgyzstan
Cape Verde
Honduras
India
Papua New Guinea
Zambia
Philippines
Nicaragua
Bangladesh
Nepal
Ghana
Lesotho
Mali
Senegal
Sierra Leone
Kenya
Madagascar
Zimbabwe
Benin
Solomon Islands
Comoros
Guinea
Mozambique
Niger
Burundi
Liberia
Malawi

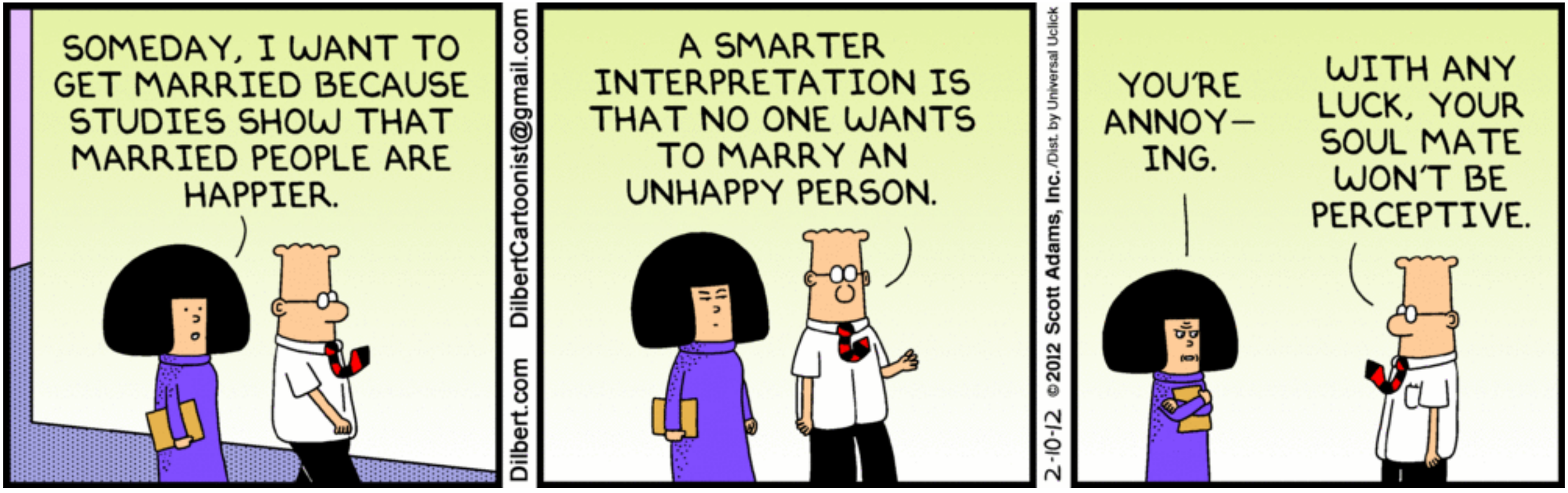
−10
−5
0
5
10

Level of Democracy
(−10 = Least Democratic, 10 = Most Democratic)



Gross National Income Per Capita
100,000
50,000
10,000
5,000
1,000
−10
−5
0
5
10
Level of Democracy
(−10 = Least Democratic, 10 = Most Democratic)
Qatar
Kuwait
Singapore
Saudi Arabia
Oman
Bahrain
Equatorial Guinea
Libya
Cuba
Belarus
Kazakhstan
Azerbaijan
Turkmenistan
China
Egypt
Jordan
Fiji
Morocco
Angola
Swaziland
Uzbekistan
Cameroon
Tajikistan
Gambia
Rwanda
Chad
Haiti
Burkina Faso
Eritrea
Ethiopia
Uganda
Togo
Central African Republic
Norway
Luxembourg
United States
Switzerland
Austria
Germany
Denmark
Netherlands
Canada
Australia
Belgium
United Kingdom
France
Italy
Sweden
Czech Republic
New Zealand
Finland
Japan
Ireland
Malaysia
Trinidad and Tobago
Israel
Slovenia
Spain
Cyprus
Latvia
Croatia
Estonia
Argentina
Portugal
Mexico
Turkey
Poland
Gabon
Botswana
Greece
Lebanon
Brazil
Romania
Hungary
Lithuania
Iraq
Suriname
Thailand
Montenegro
Panama
Bulgaria
Chile
Algeria
Ecuador
Tunisia
Colombia
Serbia
Peru
South Africa
Sri Lanka
Namibia
Dominican Republic
Uruguay
Armenia
Ukraine
Georgia
Indonesia
El Salvador
Mauritius
Bhutan
Guyana
Guatemala
Albania
Costa Rica
Nigeria
Jamaica
Pakistan
Paraguay
Mongolia
Yemen
Djibouti
Kyrgyzstan
Honduras
Cape Verde
Cambodia
India
Papua New Guinea
Zambia
Philippines
Nicaragua
Bangladesh
Nepal
Ghana
Lesotho
Mali
Senegal
Madagascar
Zimbabwe
Sierra Leone
Kenya
Guinea−Bissau
Guinea
Benin
Solomon Islands
Comoros
Mozambique
Niger
Burundi
Liberia
Malawi

correlation $\nrightarrow$ correlation



Four Ways

to Get a Correlation

causation

causation

causation

causation

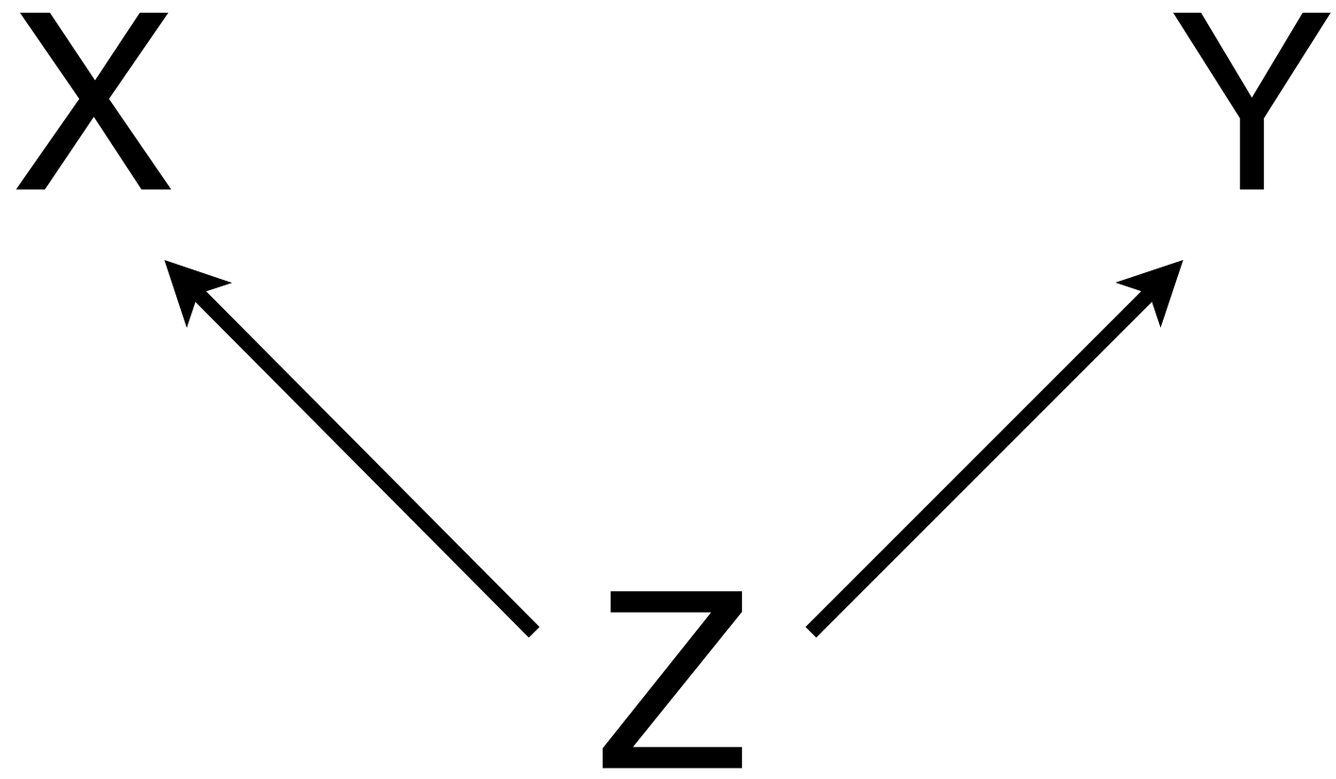
key explanatory variable

outcome variable

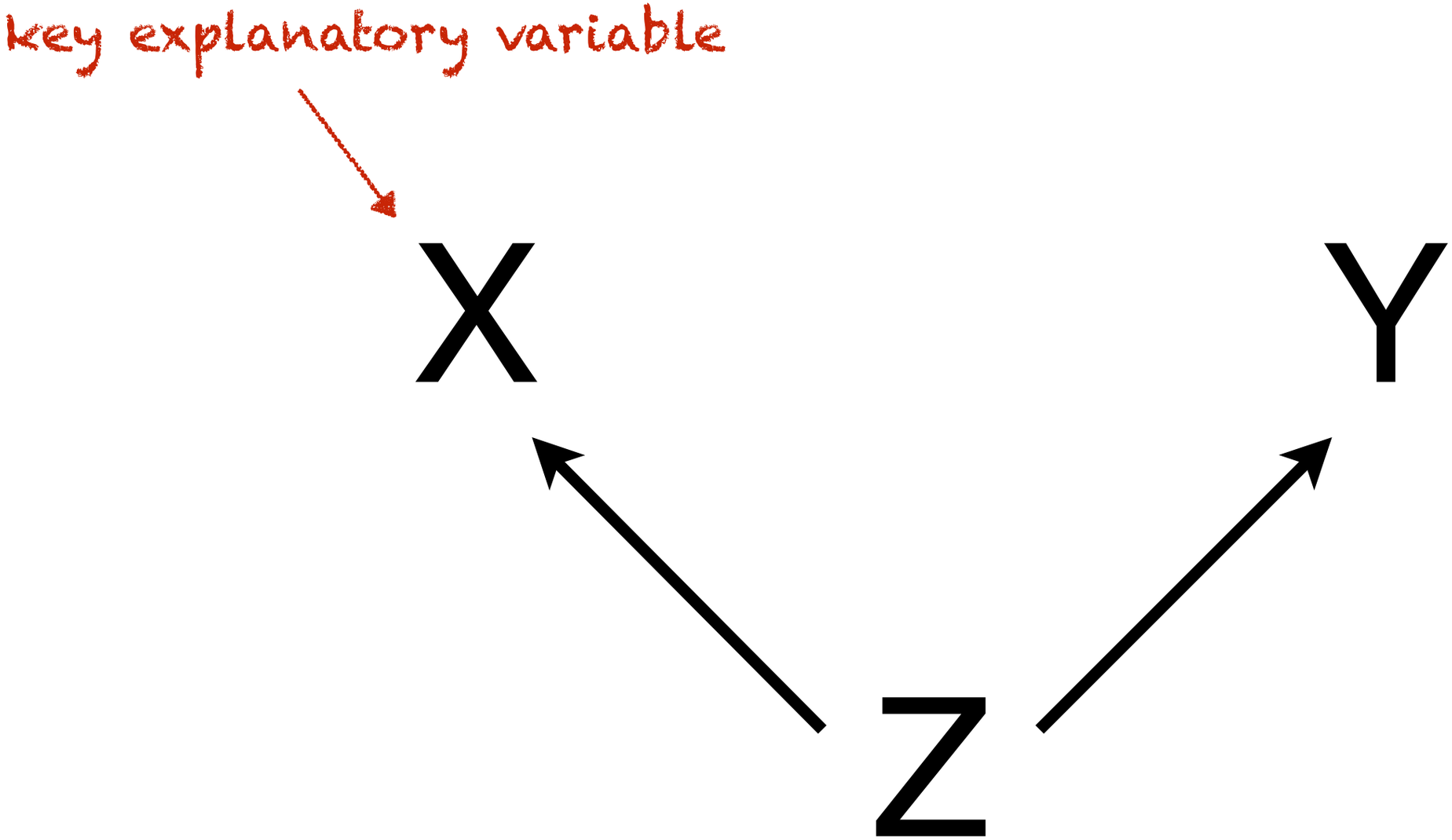
causation

spuriousness

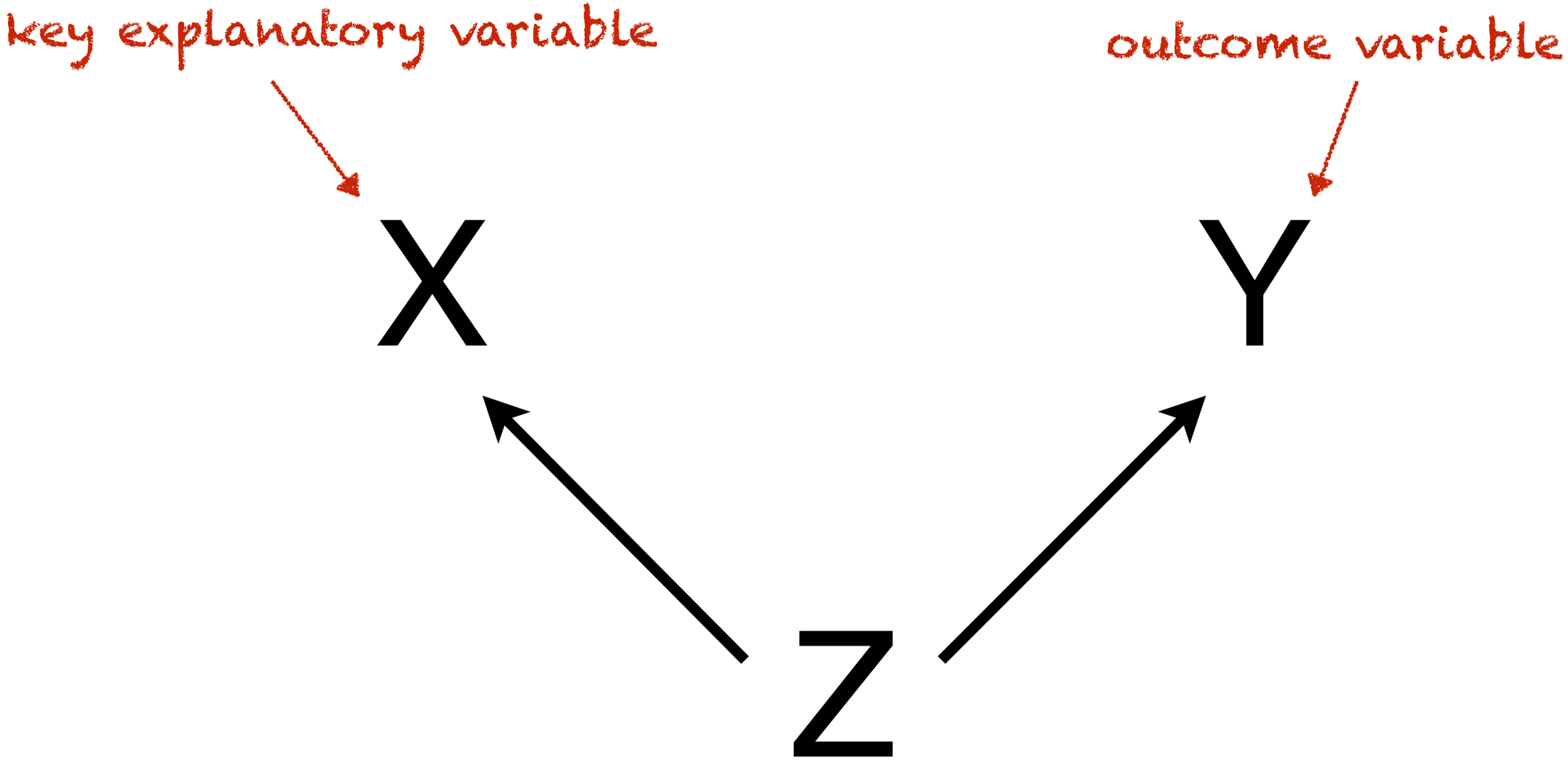
spuriousness



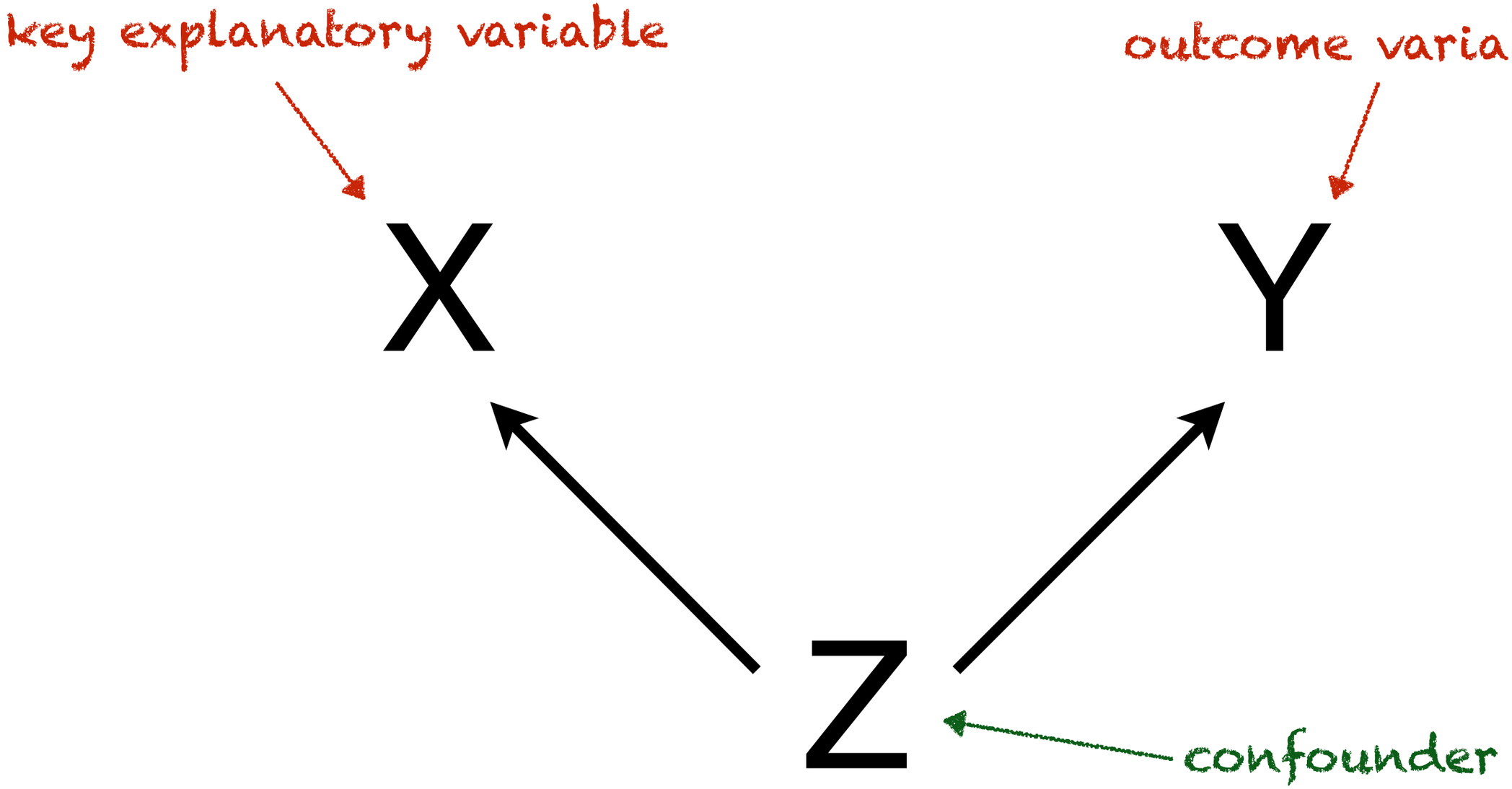
spuriousness



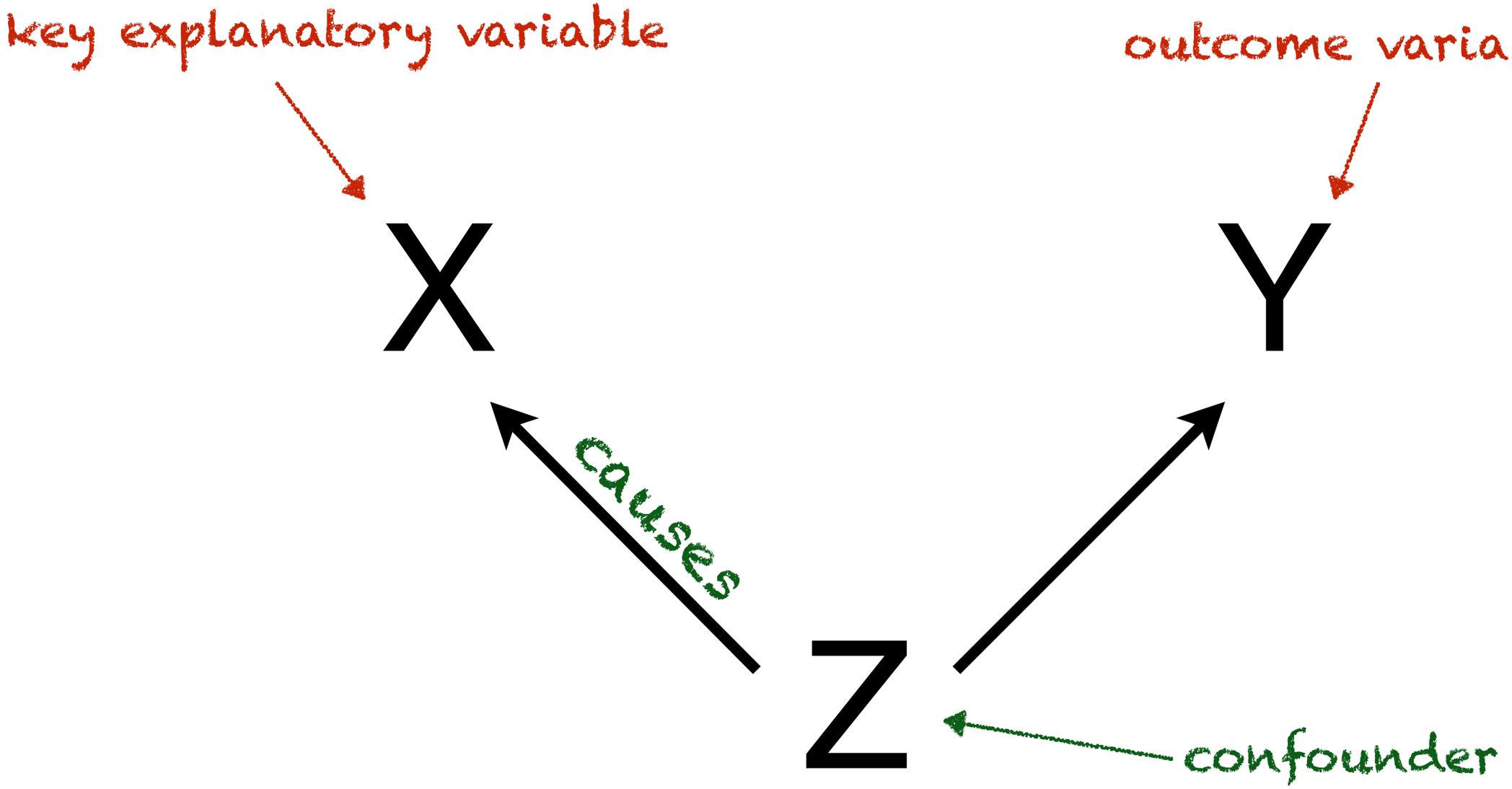
spuriousness



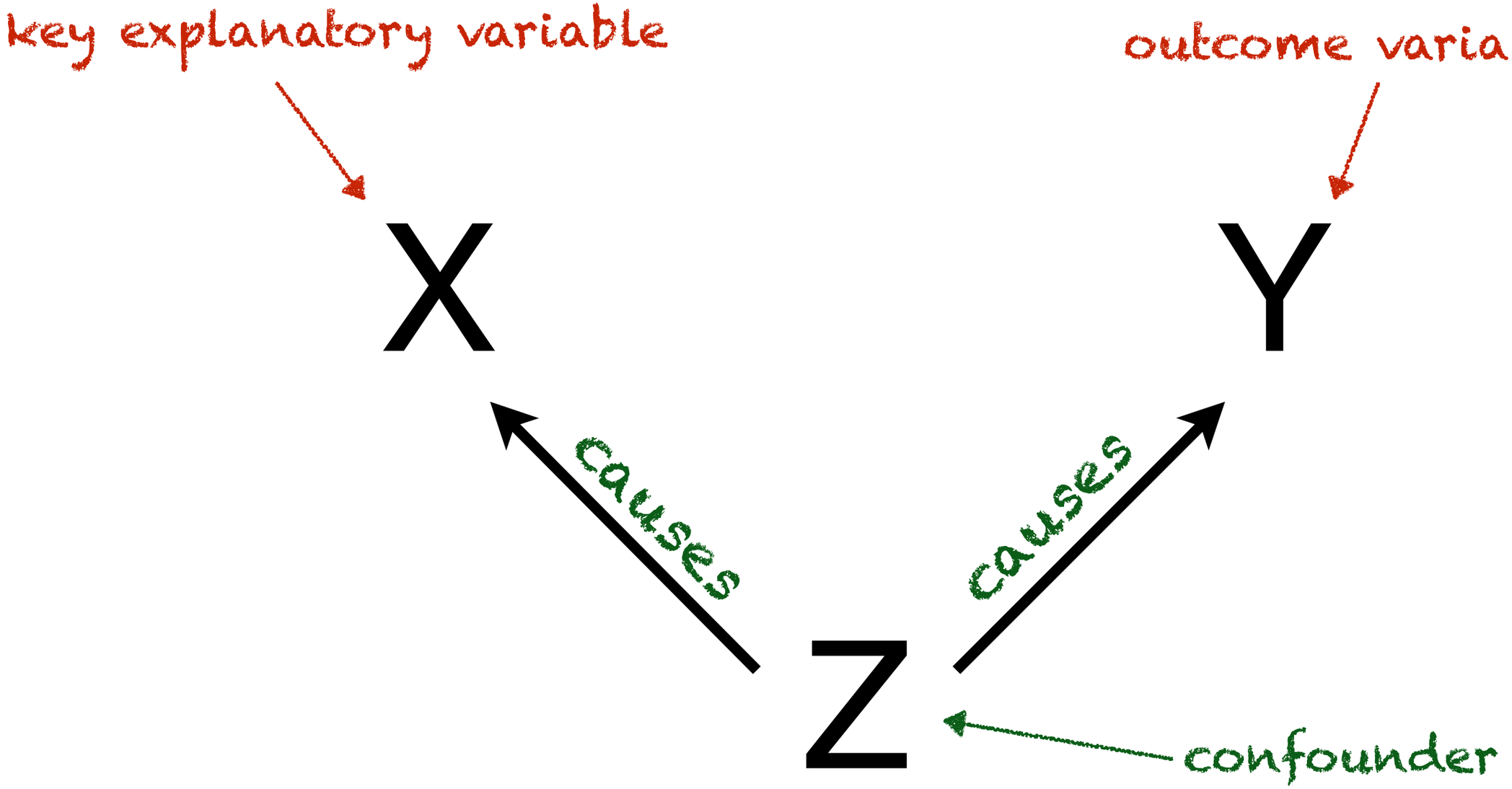
spuriousness



spuriousness

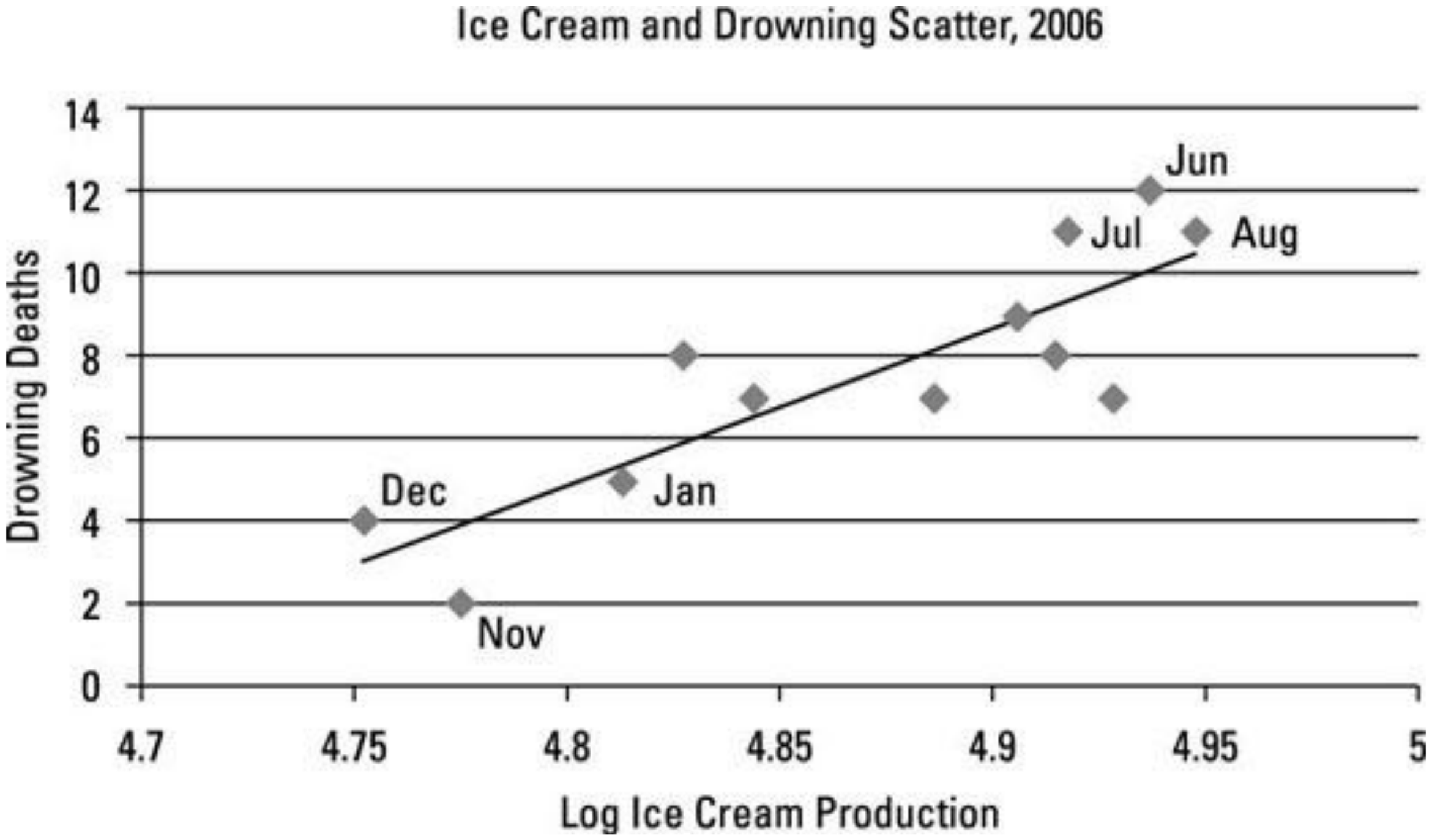


spuriousness



Note: a confounder is a variable that causes both X and Y.

spuriousness

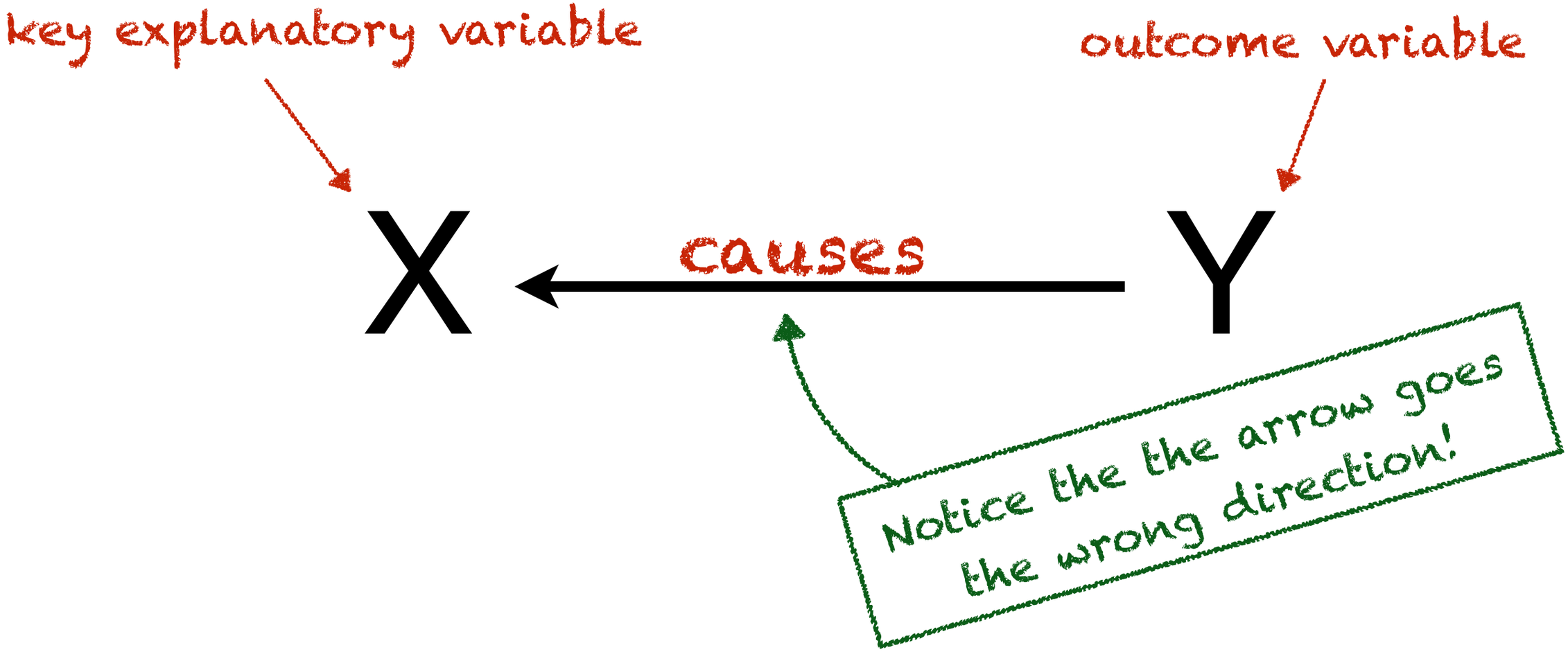

Ice Cream and Drowning Scatter, 2006
Drowning Deaths
14
12
10
8
6
4
2
0
Jun
Jul
Aug
Dec
Jan
Nov
4.7
4.75
4.8
4.85
4.9
4.95
5
Log Ice Cream Production

reverse causation

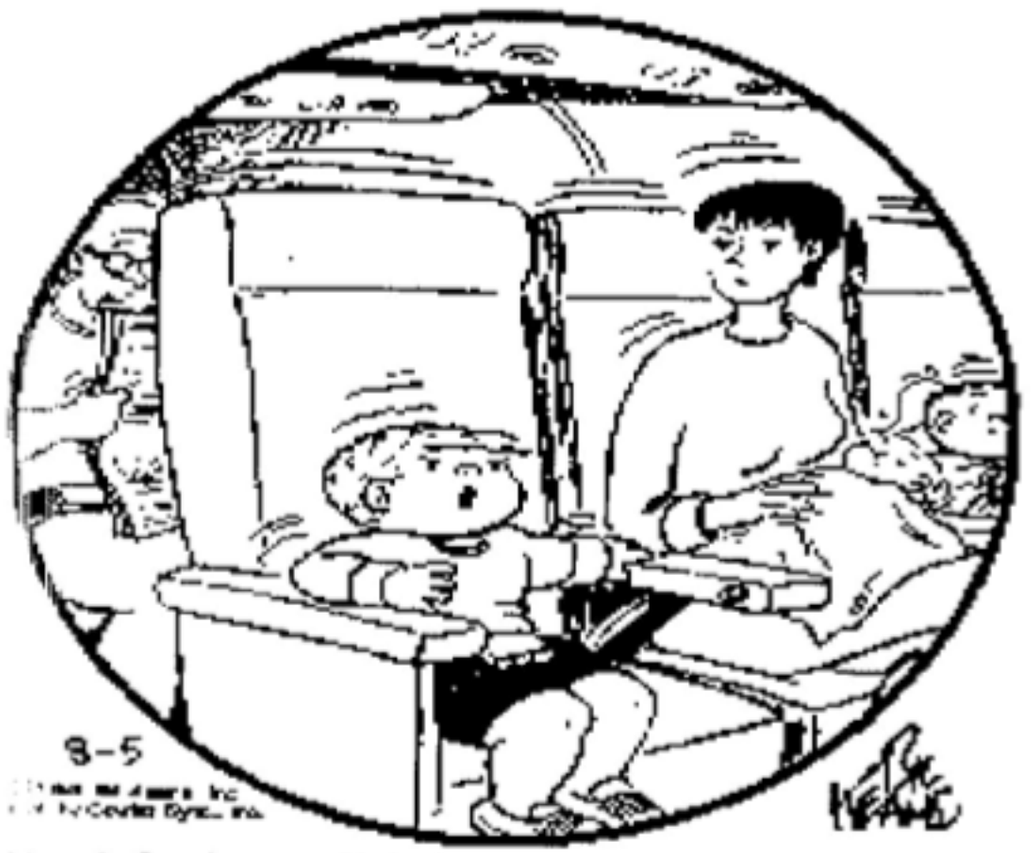
reverse causation

reverse causation

reverse causation



THE FAMILY CIRCUS



"I wish they didn't turn on that seatbelt sign so much! Every time they do, it gets bumpy."

chance

chance

X

Y

chance

key explanatory variable

X

outcome variable

Y

Sometimes, X and Y will be correlated just by chance,
even when there is no systematic relationship between the two.

chance

key explanatory variable

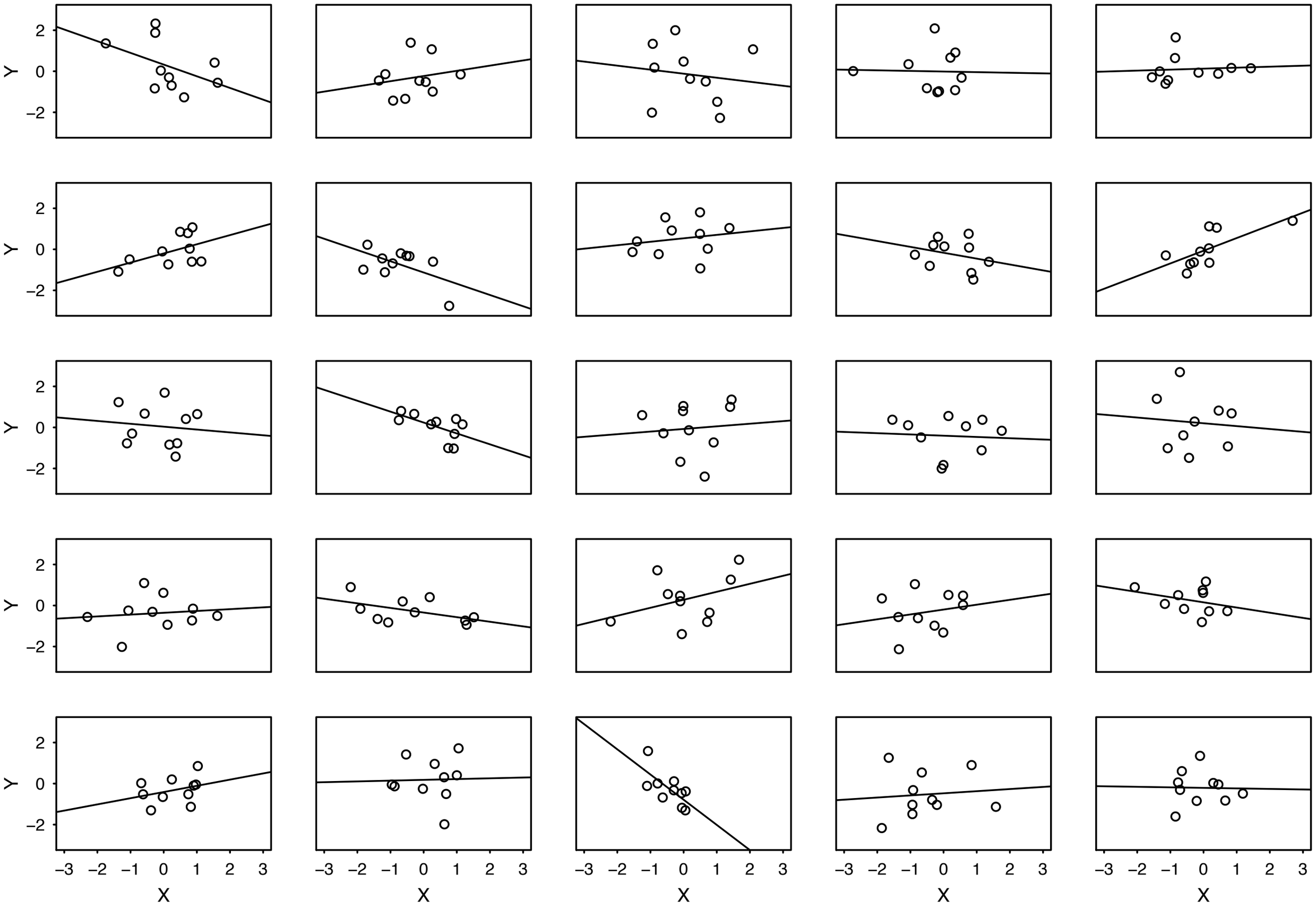
X

outcome variable

Y

Notice there is no causal arrow!

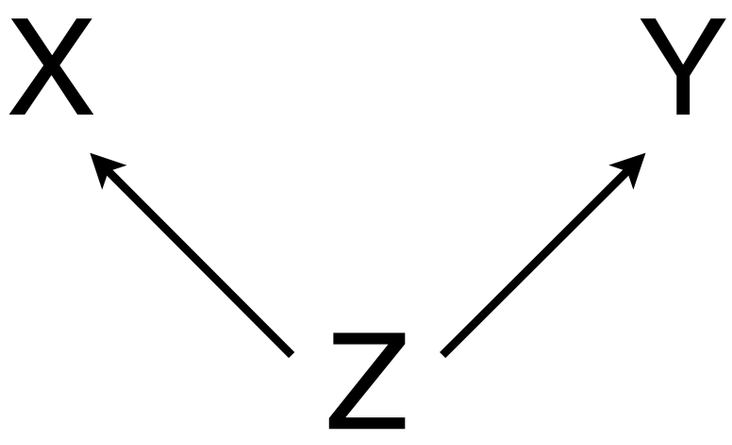
Pure Noise Generated by a Computer



causation

X ⟶ Y

spuriousness



reverse causation

X ⟵ Y

chance

no systematic relationship;
correlation simply due to chance

Ruling Out the Alternatives

spuriousness and reverse causation

- a compelling theoretical model
- randomization

spuriousness

- controlling for confounders

chance

- box model and statistical inference

Ruling Out the Alternatives

spuriousness and reverse causation

- a compelling theoretical model
- randomization

spuriousness

- controlling for confounders

chance

- box model and statistical inference

save for later...

Ruling Out the Alternatives

spuriousness and reverse causation

- a compelling theoretical model
- randomization

Discuss Today

spuriousness

- controlling for confounders

chance

- box model and statistical inference

save for later...

A Compelling Theoretical Model

A Compelling Theoretical Model

Simply explain why spuriousness and reverse causation make little theoretical sense.

A Compelling Theoretical Model

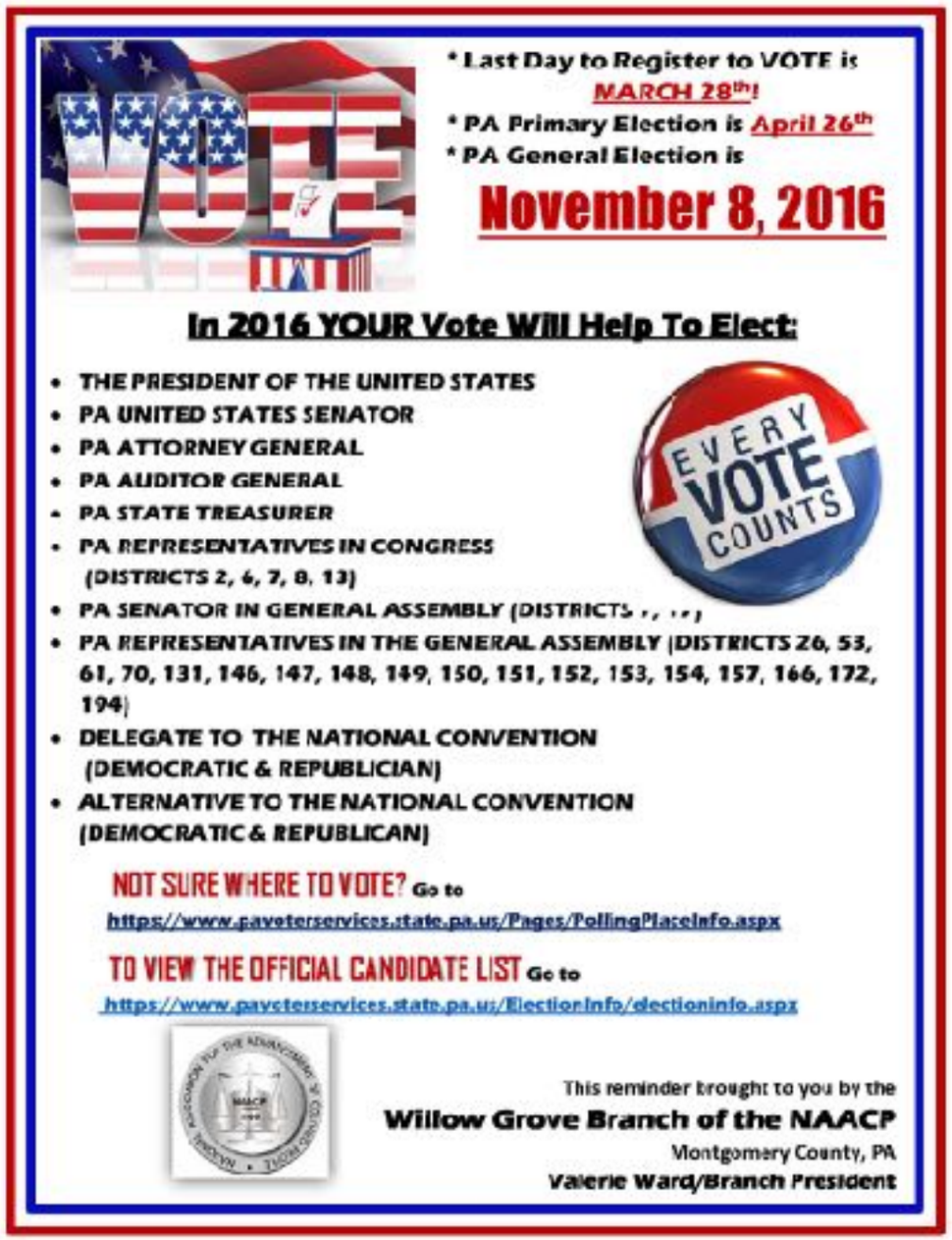
Simply explain why spuriousness and reverse causation make little theoretical sense.

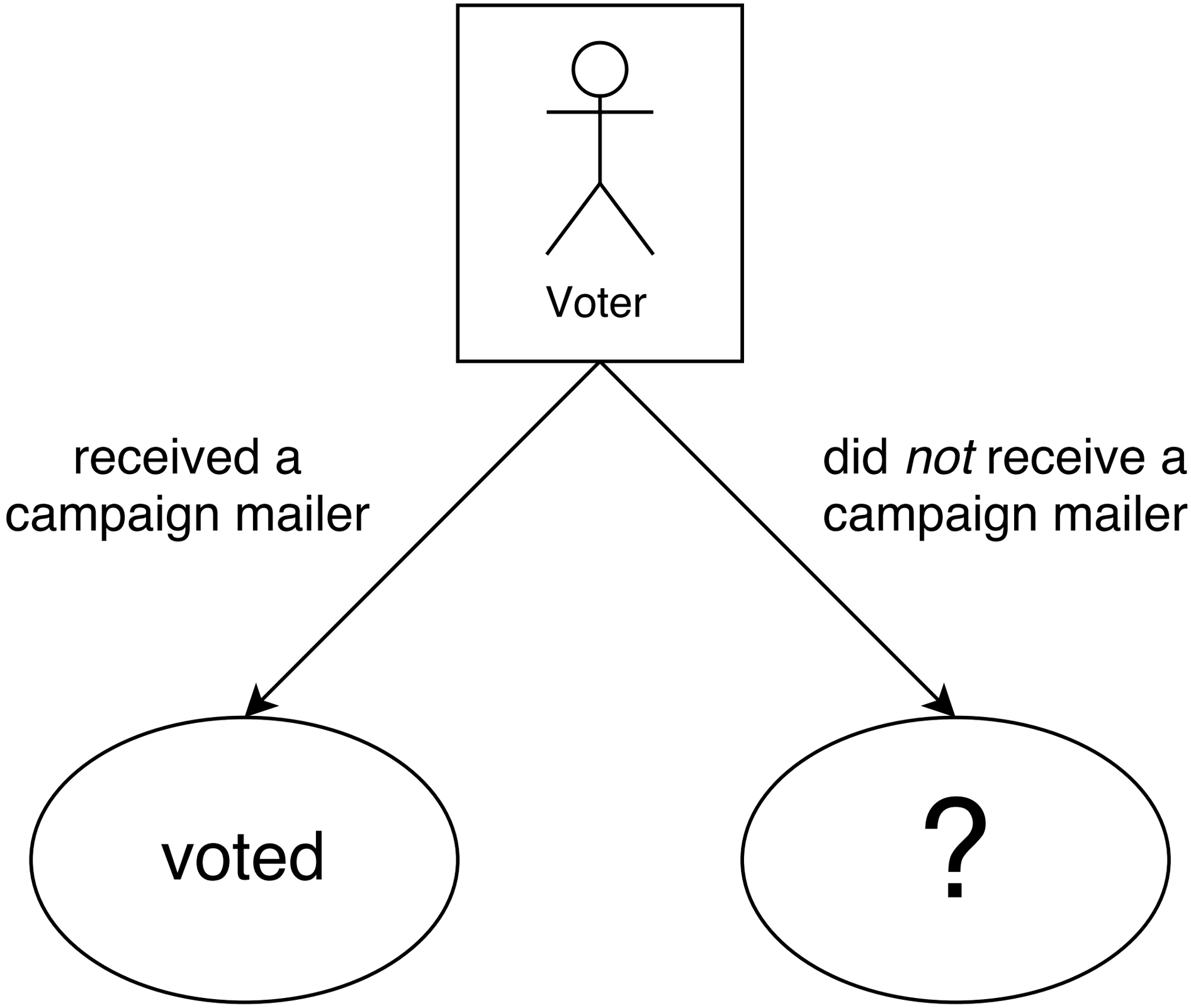
- Could it be that some other variable causes both democratic institutions and GNI? What might this be?

- Could it be that GNI causes democratic institutions?

Randomization

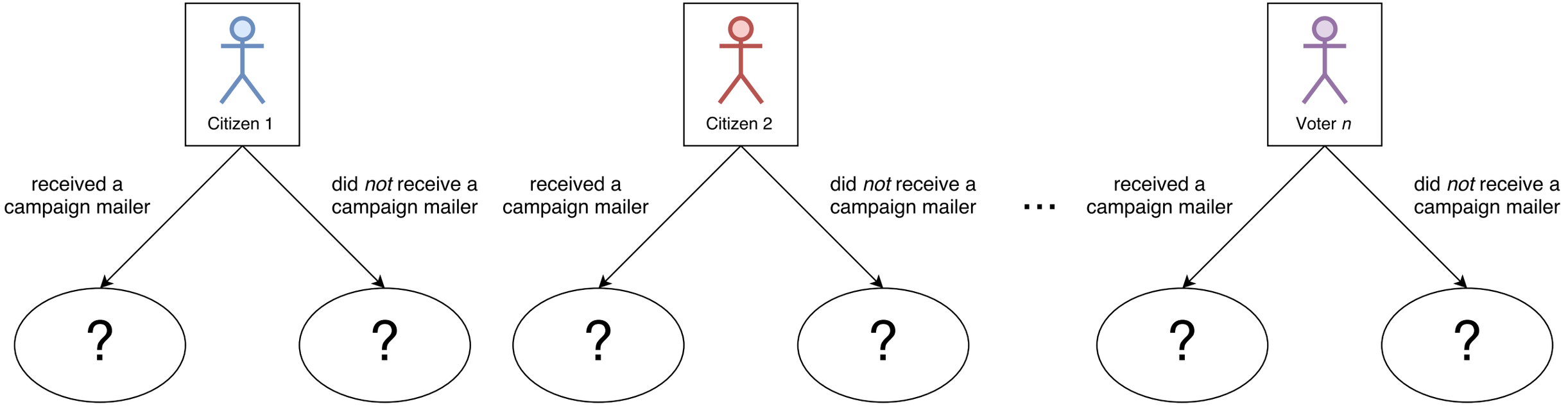
What is the effect of a campaign mailer on a citizen's decision to turn out and vote?




Voter
received a campaign mailer
did *not* receive a campaign mailer
voted
?

Imagine we're in the following **ideal situation**:

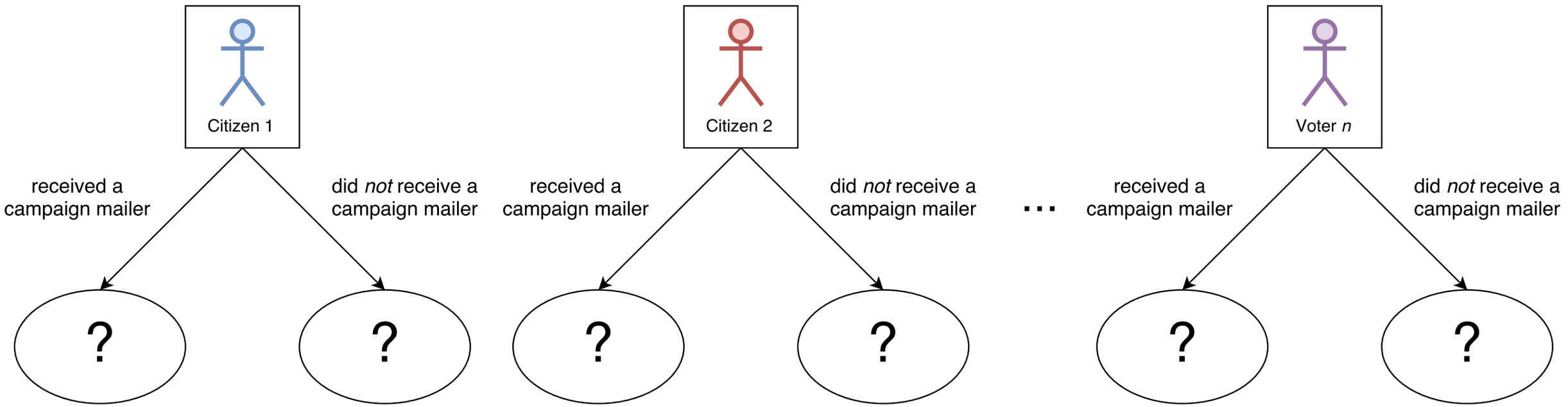
A. we have *n* potential voters,
B. the election hasn't yet happened, and
C. we can control the assignment of the treatment.



R_T^{hyp}: The hypothetical turnout Rate if everyone was in the Treatment group.

R_C^{hyp}: The hypothetical turnout Rate if everyone was in the Control group.

$R_T^{hyp} - R_C^{hyp}$: average treatment effect (ATE)


Citizen 1
received a campaign mailer
did *not* receive a campaign mailer
?
?
Citizen 2
received a campaign mailer
did *not* receive a campaign mailer
?
?
Voter *n*
received a campaign mailer
did *not* receive a campaign mailer
?
?

R_T^{obs}: The observed turnout Rate in the Treatment group.

R_C^{obs}: The observed turnout Rate in the Control group.

$$R_T^{obs} - R_C^{obs} \approx R_T^{hyp} - R_C^{hyp}$$

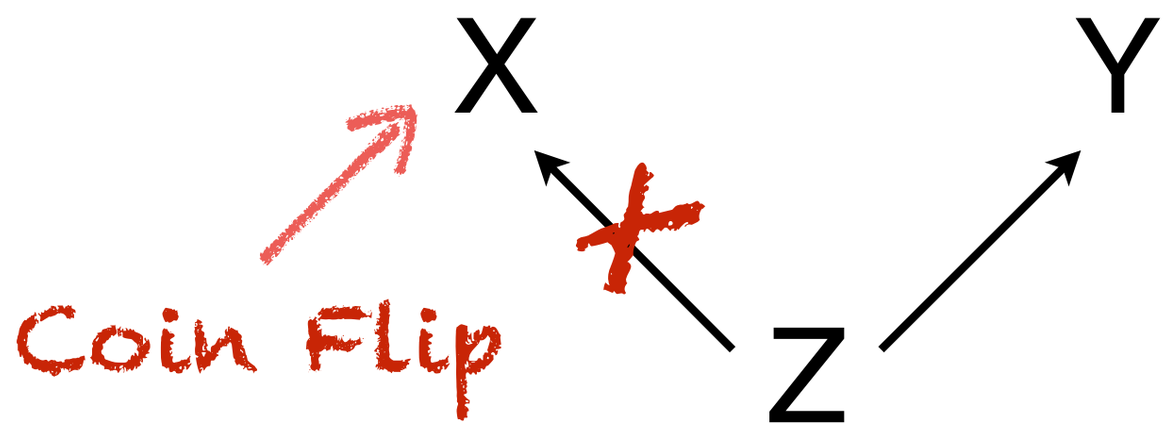
R_T^{obs}: The observed turnout Rate in the Treatment group.

R_C^{obs}: The observed turnout Rate in the Control group.

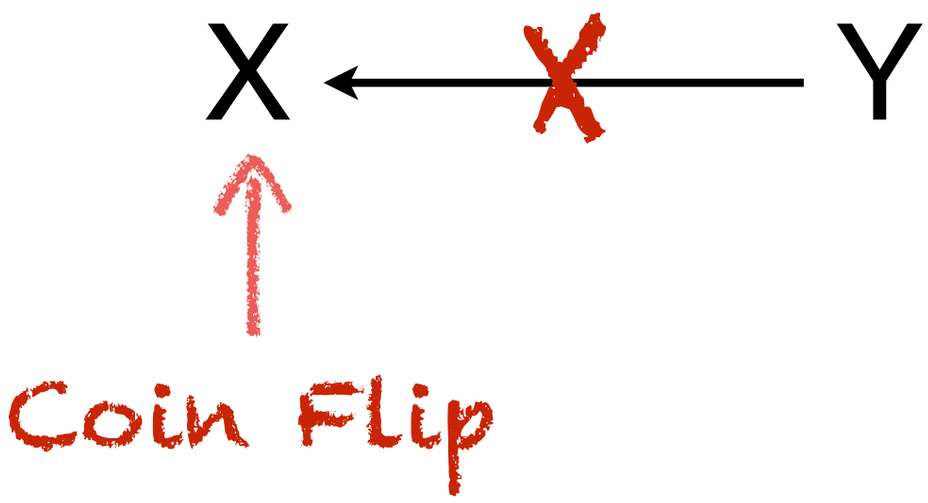
$$\underbrace{R_T^{obs} - R_C^{obs}}_{\text{estimate}} \approx \overbrace{R_T^{hyp} - R_C^{hyp}}^{\text{ATE}}$$

causation

spuriousness



reverse causation



chance

no systematic relationship;
correlation simply due to chance

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting					
N of Individuals					

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%				
N of Individuals	191,243				

Dear Registered Voter:

DO YOUR CIVIC DUTY AND VOTE!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

The whole point of democracy is that citizens are active participants in government; that we have a voice in government. Your voice starts with your vote. On August 8, remember your rights and responsibilities as a citizen. Remember to vote.

DO YOUR CIVIC DUTY — VOTE!

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%			
N of Individuals	191,243	38,218			

Dear Registered Voter:

YOU ARE BEING STUDIED!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

This year, we're trying to figure out why people do or do not vote. We'll be studying voter turnout in the August 8 primary election.

Our analysis will be based on public records, so you will not be contacted again or disturbed in any way. Anything we learn about your voting or not voting will remain confidential and will not be disclosed to anyone else.

DO YOUR CIVIC DUTY — VOTE!

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%		
N of Individuals	191,243	38,218	38,204		

Dear Registered Voter:

WHO VOTES IS PUBLIC INFORMATION!

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse.

This year, we're taking a different approach. We are reminding people that who votes is a matter of public record.

The chart shows your name from the list of registered voters, showing past votes, as well as an empty box which we will fill in to show whether you vote in the August 8 primary election. We intend to mail you an updated chart when we have that information.

We will leave the box blank if you do not vote.

DO YOUR CIVIC DUTY—VOTE!

OAK ST	Aug 04	Nov 04	Aug 06
9999 ROBERT WAYNE		Voted	_____
9999 LAURA WAYNE	Voted	Voted	_____

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	
N of Individuals	191,243	38,218	38,204	38,218	

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	_____
9995 JENNIFER KAY SMITH		Voted	_____
9997 RICHARD B JACKSON		Voted	_____
9999 KATHY MARIE JACKSON		Voted	_____
9999 BRIAN JOSEPH JACKSON		Voted	_____
9991 JENNIFER KAY THOMPSON		Voted	_____
9991 BOB R THOMPSON		Voted	_____
9993 BILL S SMITH			_____
9989 WILLIAM LUKE CASPER		Voted	_____
9989 JENNIFER SUE CASPER		Voted	_____
9987 MARIA S JOHNSON	Voted	Voted	_____
9987 TOM JACK JOHNSON	Voted	Voted	_____
9987 RICHARD TOM JOHNSON		Voted	_____
9985 ROSEMARY S SUE		Voted	_____
9985 KATHRYN L SUE		Voted	_____
9985 HOWARD BEN SUE		Voted	_____
9983 NATHAN CHAD BERG		Voted	_____
9983 CARRIE ANN BERG		Voted	_____
9981 EARL JOEL SMITH			_____
9979 DEBORAH KAY WAYNE		Voted	_____
9979 JOEL R WAYNE		Voted	_____

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

Review Exercises

1. I write that, under randomization, $R_T^{obs} - R_C^{obs} \approx R_T^{hyp} - R_C^{hyp}$

   A. What do each of these four quantities refer to? What do we call the left-hand side? The right-hand side?

   B. Notice that the equality is not exact. Instead, it is approximate. What is the only reason it is not exact?

   C. Explain why randomization allows us to rule out spuriousness and reverse causation.

2. Describe the design of Gerber and Green's turnout experiment. Describe the results (i.e., what percent of each group voted?). Discuss whether you can rule out any of the four possible ways to obtain a correlation.