Causation

If we have high correlation, we'd like to determine causation.

Causation

If we have high correlation, we'd like to determine causation.

To visually represent the direction of causality between variables, use arrows. For example, if x causes y, we draw an arrow from x to y.

Causation

If we have high correlation, we'd like to determine causation.

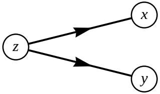
To visually represent the direction of causality between variables, use arrows. For example, if x causes y, we draw an arrow from x to y.

The ways in which two variables may have strong correlation are:

I. Simple Causality 

II. Reverse Causality 

III. Mutual Causality 

IV. Hidden/Confounding Variable 

V. Complete Accident/Coincidence 

Simple Causality

I. Simple Causality 

We say that variables x and y are related by **simple causality** if the level of x *determines* the level of y.

Simple Causality

I. Simple Causality 

We say that variables x and y are related by **simple causality** if the level of x *determines* the level of y.

Example 2 (pp. 171–173): High blood pressure. There is high correlation in the plot of blood pressure vs. deaths from heart disease.

Simple Causality

I. Simple Causality 

We say that variables x and y are related by **simple causality** if the level of x *determines* the level of y.

Example 2 (pp. 171–173): High blood pressure. There is high correlation in the plot of blood pressure vs. deaths from heart disease.

A chain of causation argues for simple causality:

high blood pressure $\rightarrow$ arteries clog $\rightarrow$
lack of oxygen in heart $\rightarrow$ heart disease

Simple Causality

I. Simple Causality 

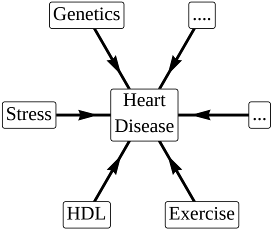
We say that variables x and y are related by **simple causality** if the level of x *determines* the level of y.

Example 2 (pp. 171–173): High blood pressure. There is high correlation in the plot of blood pressure vs. deaths from heart disease.

A chain of causation argues for simple causality:

high blood pressure $\rightarrow$ arteries clog $\rightarrow$ lack of oxygen in heart $\rightarrow$ heart disease

Many factors have been determined that increase the chance for heart disease.



Reverse Causality

II. Reverse Causality 

We say that variables x and y are related by **reverse causality** if the level of x *is determined by* the level of y.

Reverse Causality

II. Reverse Causality 

We say that variables x and y are related by **reverse causality** if the level of x *is determined by* the level of y.

Example. Islanders in South Pacific found a correlation between health and body lice.

Healthy people had body lice and sick people didn't.

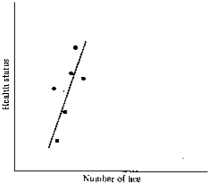


Figure 7

Reverse Causality

II. Reverse Causality 

We say that variables x and y are related by **reverse causality** if the level of x *is determined by* the level of y.

Example. Islanders in South Pacific found a correlation between health and body lice.

Healthy people had body lice and sick people didn't.

Hence: More body lice means better health.

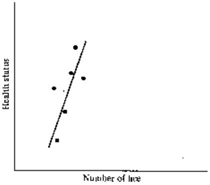


Figure 7

Reverse Causality

II. Reverse Causality 

We say that variables x and y are related by **reverse causality** if the level of x *is determined by* the level of y.

Example. Islanders in South Pacific found a correlation between health and body lice.

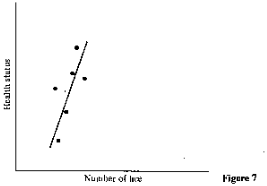


Figure 7

Healthy people had body lice and sick people didn't.

Hence: More body lice means better health.

However, everyone had lice. Lice just preferred healthy hosts.

Reverse Causality

II. Reverse Causality $x \leftarrow y$

We say that variables x and y are related by **reverse causality** if the level of x *is determined by* the level of y.

Example. Islanders in South Pacific found a correlation between health and body lice.

Healthy people had body lice and sick people didn't.

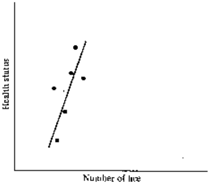


Figure 7

Hence: More body lice means better health.

However, everyone had lice. Lice just preferred healthy hosts.

Example. Human birth rate and stork population:

Storks bring babies.

Mutual Causality / Feedback

III. Mutual Causality 

We say that variables x and y are related by **mutual causality** if changes in x produce changes in y and vice versa.

Mutual Causality / Feedback

III. Mutual Causality 

We say that variables x and y are related by **mutual causality** if changes in x produce changes in y and vice versa.

Example. Car dealers.

There is a strong correlation between
Dealer car sales and Dealer advertising budget.

Mutual Causality / Feedback

III. Mutual Causality 

We say that variables x and y are related by **mutual causality** if changes in x produce changes in y and vice versa.

Example. Car dealers.

There is a strong correlation between
Dealer car sales and Dealer advertising budget.

Do car sales pay for advertising
or does advertising drive sales?

Mutual Causality / Feedback

III. Mutual Causality 

We say that variables x and y are related by **mutual causality** if changes in x produce changes in y and vice versa.

Example. Car dealers.

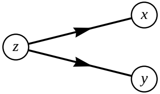
There is a strong correlation between
Dealer car sales and Dealer advertising budget.

Do car sales pay for advertising
or does advertising drive sales?

These are mutually reinforcing.
This is an example of mutual causality.

Hidden Variable Causes Both

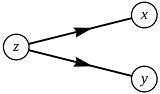
IV. Hidden/Confounding Variable



We say that x and y are in a **spurious relationship** if the levels of both x and y are determined by the level of a **confounding variable** z.

Hidden Variable Causes Both

IV. Hidden/Confounding Variable



We say that x and y are in a **spurious relationship** if the levels of both x and y are determined by the level of a **confounding variable** z.

Example. In a city, the number of churches there are is highly correlated with the number of liquor stores.

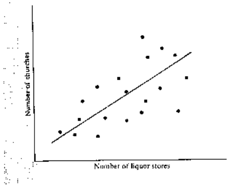
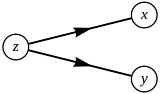


Figure 10

Hidden Variable Causes Both

IV. Hidden/Confounding Variable



We say that x and y are in a **spurious relationship** if the levels of both x and y are determined by the level of a **confounding variable** z.

Example. In a city, the number of churches there are is highly correlated with the number of liquor stores.

- Simple causation would imply:

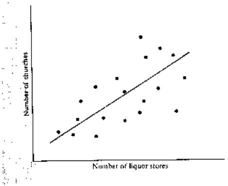
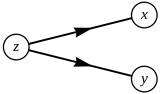


Figure 10

Hidden Variable Causes Both

IV. Hidden/Confounding Variable



We say that x and y are in a **spurious relationship** if the levels of both x and y are determined by the level of a **confounding variable** z.

Example. In a city, the number of churches there are is highly correlated with the number of liquor stores.

- Simple causation would imply:
- Reverse causation would imply:

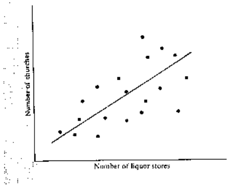
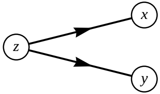


Figure 10

Hidden Variable Causes Both

IV. Hidden/Confounding Variable



We say that x and y are in a **spurious relationship** if the levels of both x and y are determined by the level of a **confounding variable** z.

Example. In a city, the number of churches there are is highly correlated with the number of liquor stores.

- Simple causation would imply:
- Reverse causation would imply:

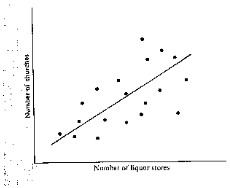


Figure 10

In this instance, there is a confounding variable: ______________________.

Complete Accident

V. Complete Accident/Coincidence 

If none of the above four cases apply, x and y are unrelated.

Complete Accident

V. Complete Accident/Coincidence 

If none of the above four cases apply, x and y are unrelated.

Take two dice. Roll each five times. Plot the value of one die versus the value of the other die for the five rolls. Often there will be no correlation.

Complete Accident

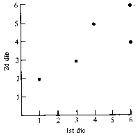
V. Complete Accident/Coincidence (x) (y)

If none of the above four cases apply, x and y are unrelated.

Take two dice. Roll each five times. Plot the value of one die versus the value of the other die for the five rolls. Often there will be no correlation.

One instance of correlation occurred, with an R^2 of 0.672 (relatively high!)

Roll	Number showing 1st die	2d die
#1	1	2
#2	3	3
#3	4	5
#4	5	4
#5	6	6



Complete Accident

V. Complete Accident/Coincidence x y

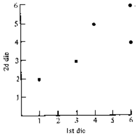
If none of the above four cases apply, x and y are unrelated.

Take two dice. Roll each five times. Plot the value of one die versus the value of the other die for the five rolls. Often there will be no correlation.

One instance of correlation occurred, with an R^2 of 0.672 (relatively high!)

An example of a correlation by coincidence.

	Number showing	
Roll	1st die	2d die
#1	1	2
#2	3	3
#3	4	5
#4	5	4
#5	6	6



Complete Accident

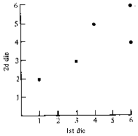
V. Complete Accident/Coincidence

If none of the above four cases apply, x and y are unrelated.

Take two dice. Roll each five times. Plot the value of one die versus the value of the other die for the five rolls. Often there will be no correlation.

One instance of correlation occurred, with an R^2 of 0.672 (relatively high!)

Roll	Number showing 1st die	2d die
#1	1	2
#2	3	3
#3	4	5
#4	5	4
#5	6	6



An example of a correlation by coincidence.

Example. Perhaps with students and SSN's?

Complete Accident

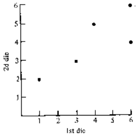
V. Complete Accident/Coincidence

If none of the above four cases apply, x and y are unrelated.

Take two dice. Roll each five times. Plot the value of one die versus the value of the other die for the five rolls. Often there will be no correlation.

One instance of correlation occurred, with an R^2 of 0.672 (relatively high!)

Roll	Number showing: 1st die	Number showing: 2d die
#1	1	2
#2	3	3
#3	4	5
#4	5	4
#5	6	6



An example of a correlation by coincidence.

Example. Perhaps with students and SSN's?

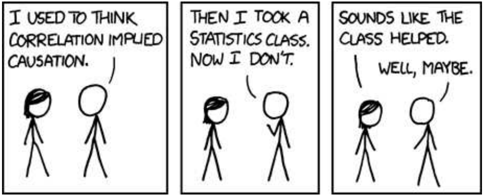
- The chance of this occurring decreases as more observations are taken.

Correlation does not imply causation!

Groupwork: Justify the correlations between the following variables:

- As ice cream sales increase, the rate of drowning deaths increase.
- The more firemen fighting the fire, the larger the fire grows.
- With fewer pirates on the open seas, global warming has increased.
- The more people in my Facebook group, the faster it grows.

What is the joke below?



Source: `http://xkcd.com/552/`