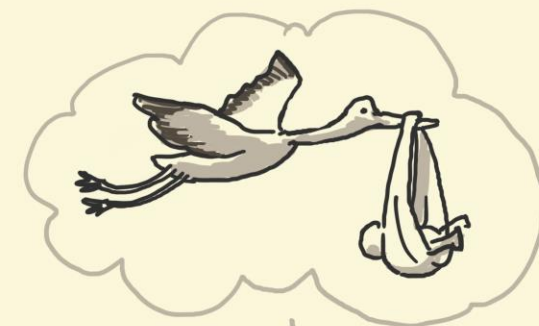
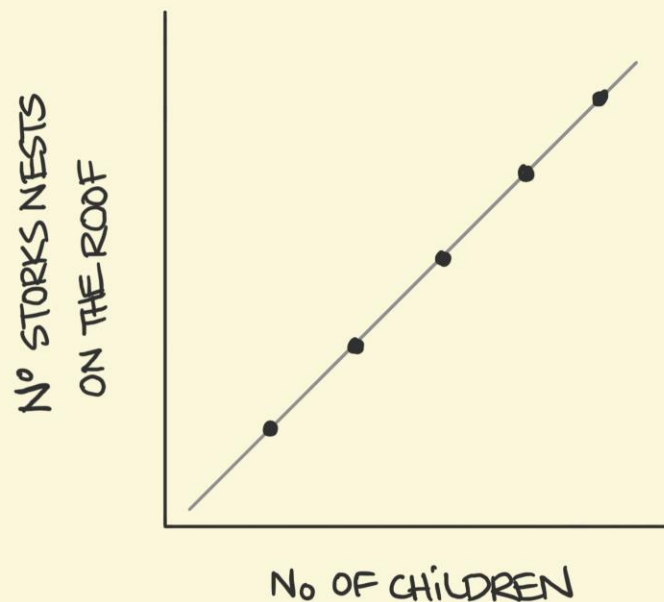


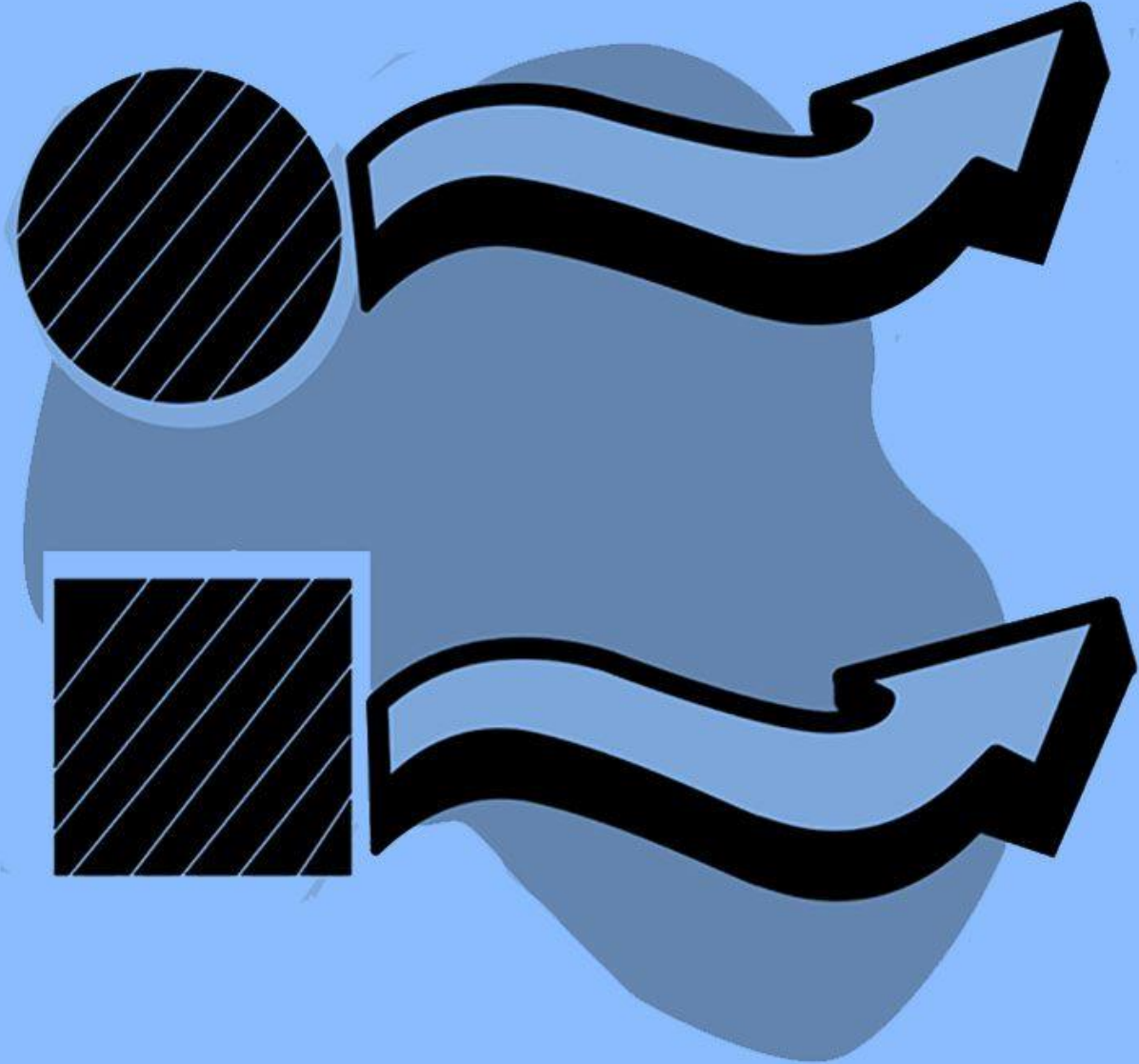
# CORRELATION IS NOT CAUSATION



sketchplanations

# Causal Relationships

Positive, Negative Correlations



# Correlation

[, kor-ə-'lā-shən]

Two variables move in relation to each other. They both move up OR one goes up when the other goes down.

# Pearson's linear relationship types

◆ Aka - Correlation Types OR Correlation Direction



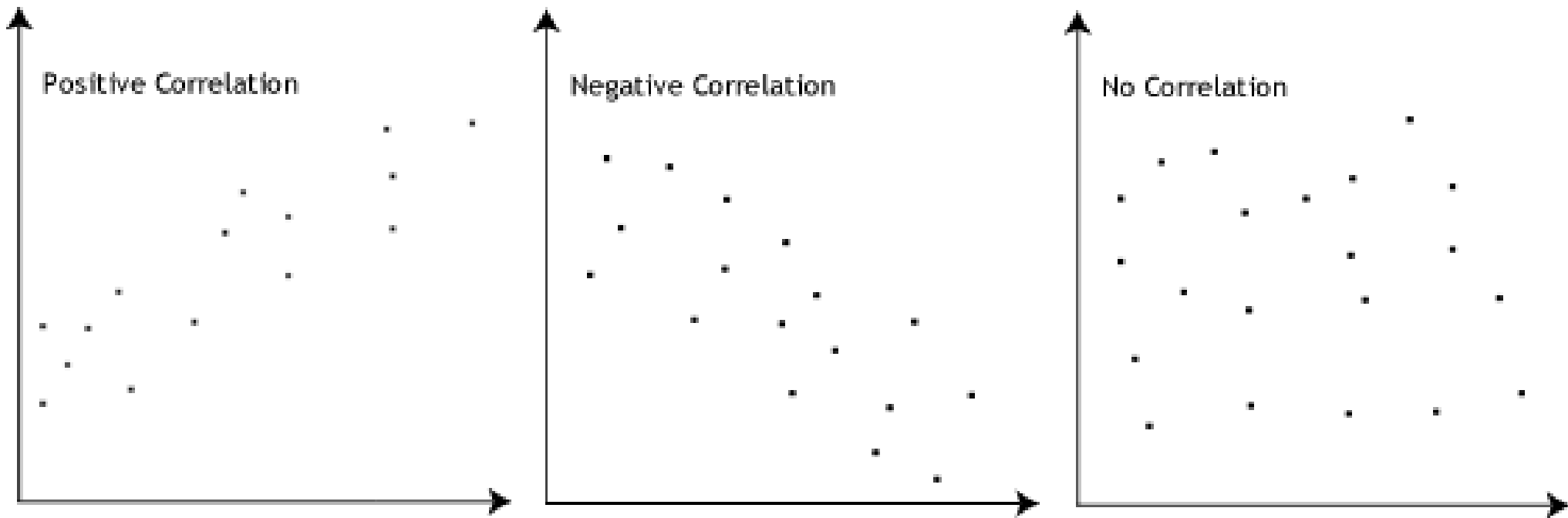
## Positive linear relationship

An example is a person's income that increases with age.



## Negative linear relationship

An example is if a vehicle increases its speed, the time it takes to move decreases, and vice versa.



- A **positive** correlation is a relationship between two variables where if one variable increases, the other one also increases.
- A **negative** correlation is a relationship between two variables where if one variable increases, the other one decreases.

Does each set of variables show a positive correlation, negative correlation, or no correlation? Explain your reasoning.

- (a) The distance Angel swims and the time she takes
- (b) The distance Angel swims and the time Dolores spends studying
- (c) The temperature and the amount of clothing people wear
- (d) The number of cats a dog sees and how often it barks
- (e) The temperature in Vancouver and the temperature in Toronto
- (f) The distance a jogger runs and the time the jogger runs





# Correlation Coefficient

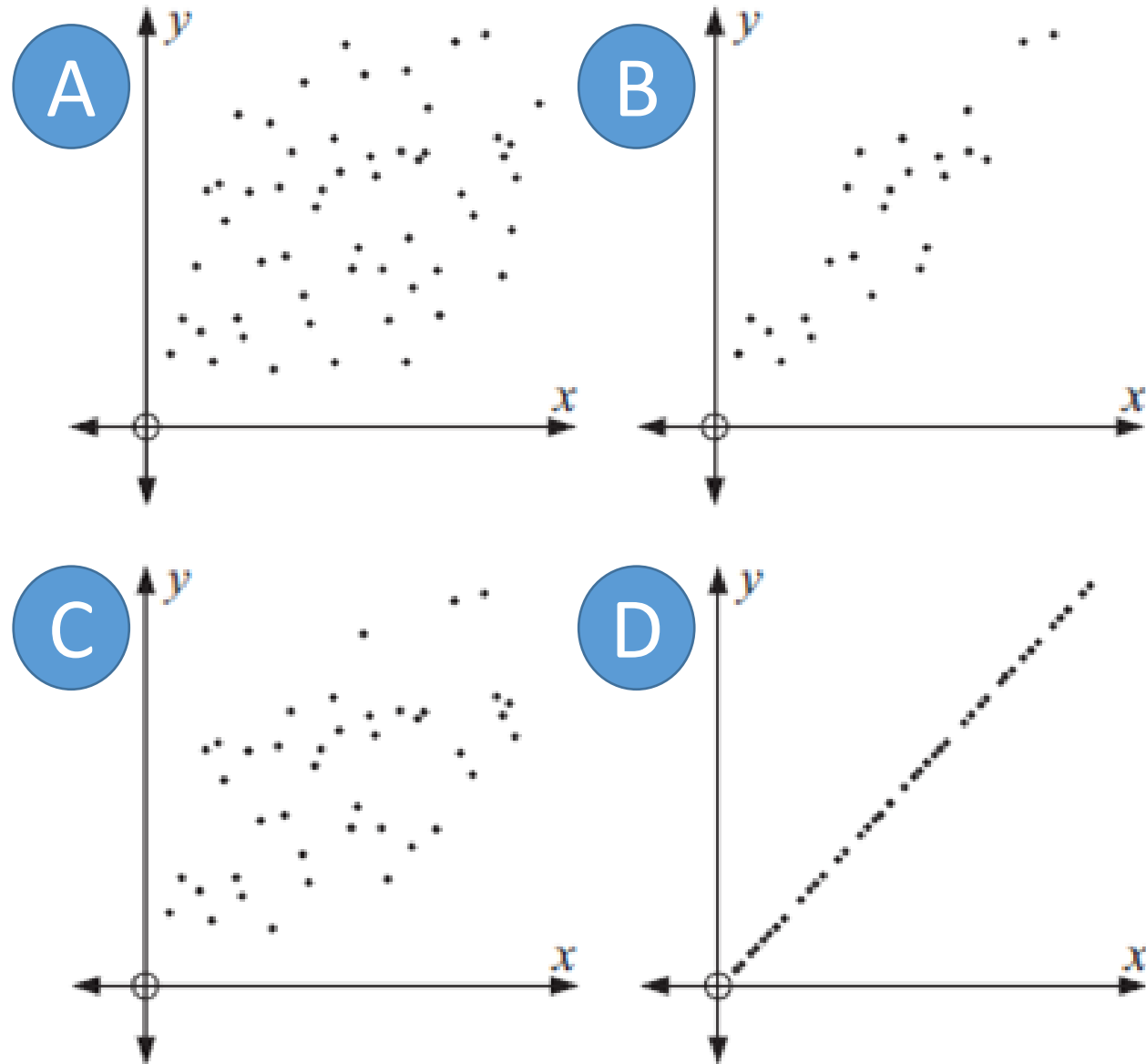
*[,kór-ə-'lā-shən ,kō-ə-'fi-shənt]*

A statistical measure of the strength of the relationship between the relative movements of two variables.

*Put these graphs in order from weakest to greatest strength.*

## Strength of a Relationship

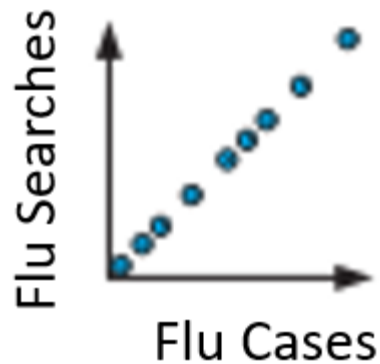
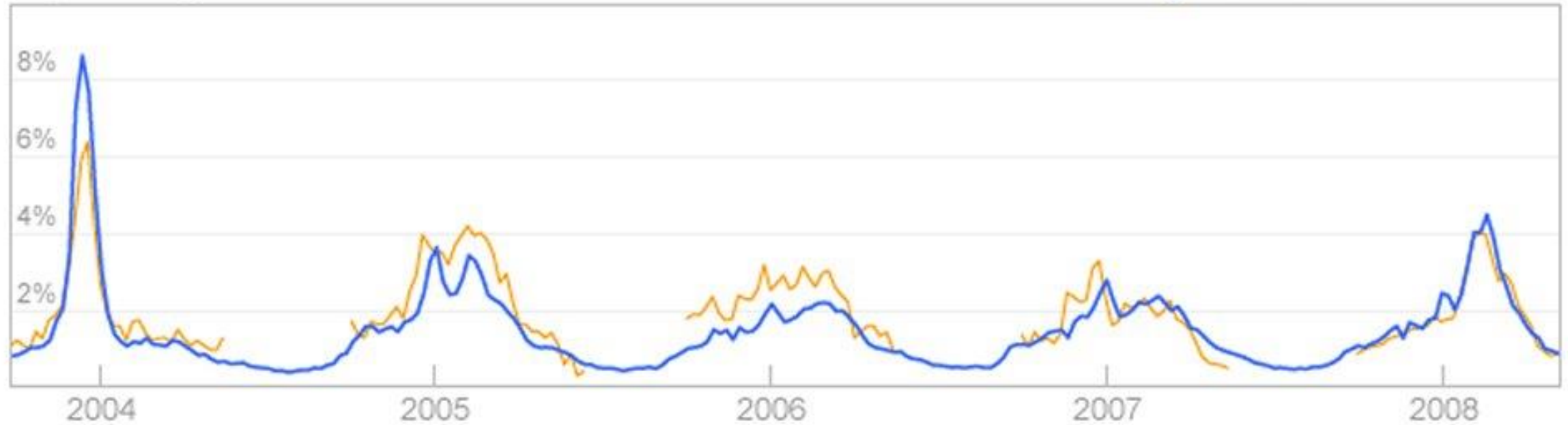
- A relationship between two variables is considered strong if the data is closely grouped around a line of best fit.



# Annual U.S. Flu Activity - Mid-Atlantic Region

ILI percentage

● Google Flu Trends ● CDC Data



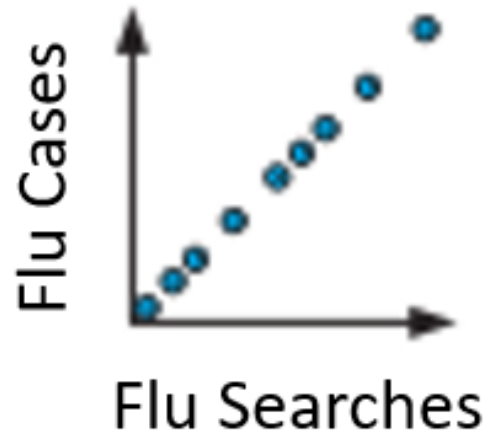
As the number of flu cases (X) increases, the number of flu searches (Y) also increases.

They are **positively** correlated.

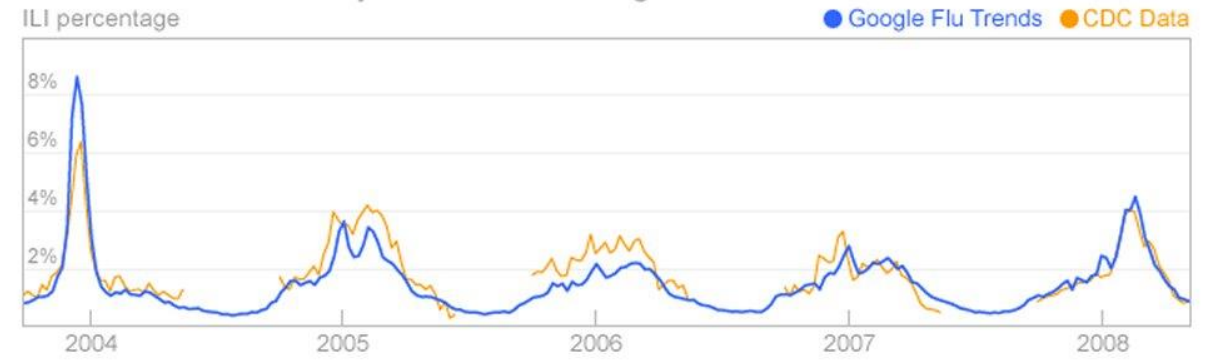


When two things are correlated,  
one of three things is happening:

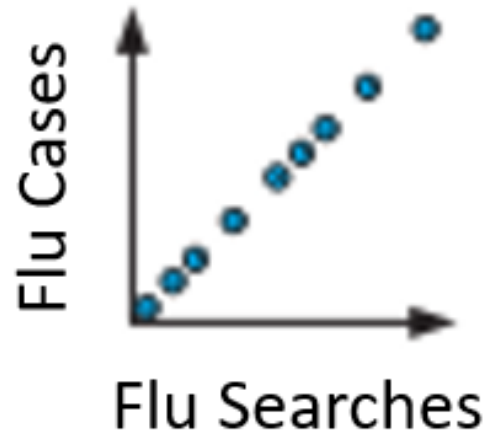
1.  $X$  causes  $Y$  (causation)
2.  $Y$  causes  $X$  (reverse-causation)
3.  $Z$  causes  $X$  and  $Y$  (spuriousness)



Annual U.S. Flu Activity - Mid-Atlantic Region



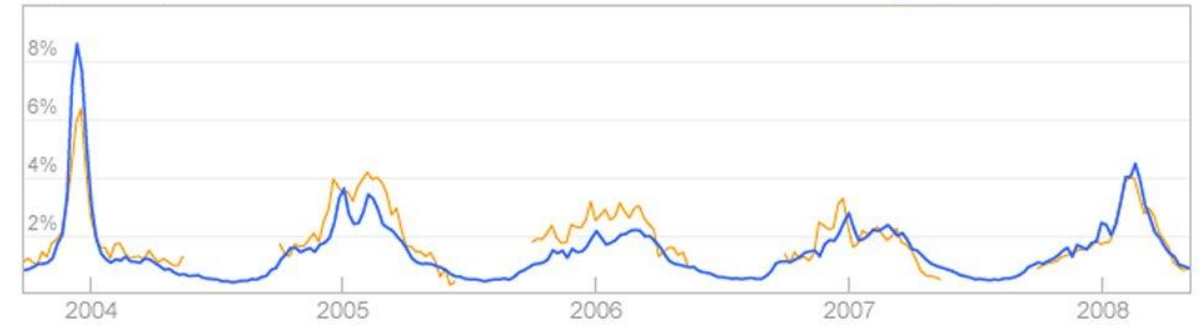
↑ flu cases (X) → ↑ flu searches (Y).  
They are **positively** correlated.



Annual U.S. Flu Activity - Mid-Atlantic Region

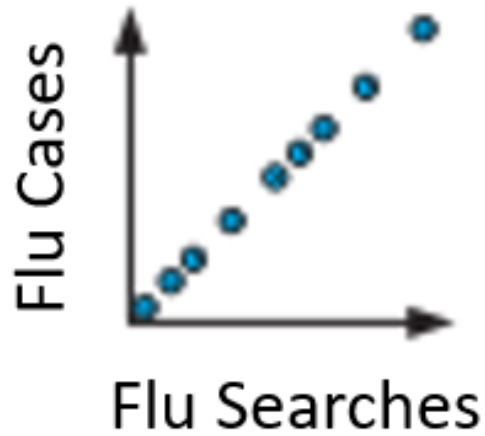
ILI percentage

● Google Flu Trends ● CDC Data

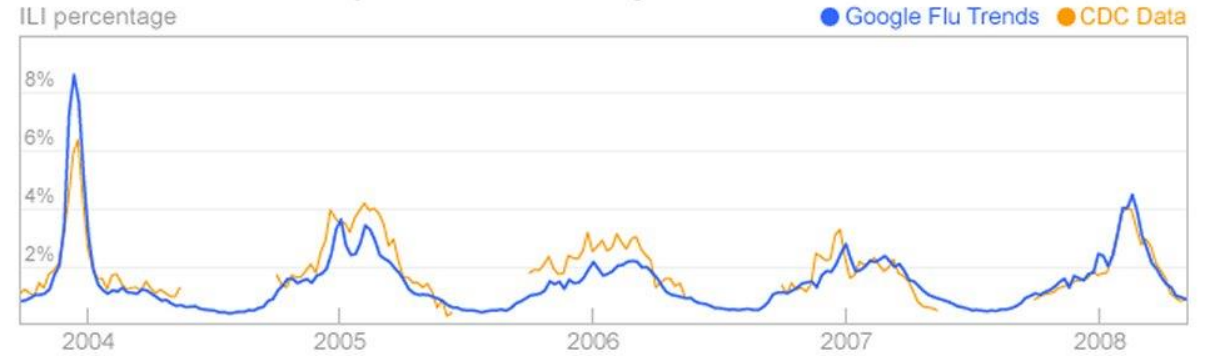


↑ flu cases (X) → ↑ flu searches (Y).  
They are **positively** correlated.

This means one of the following is happening:



Annual U.S. Flu Activity - Mid-Atlantic Region

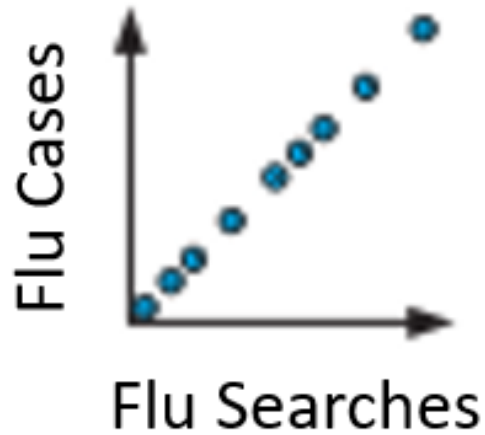


↑ flu cases (X) → ↑ flu searches (Y).

They are **positively** correlated.

This means one of the following is happening:

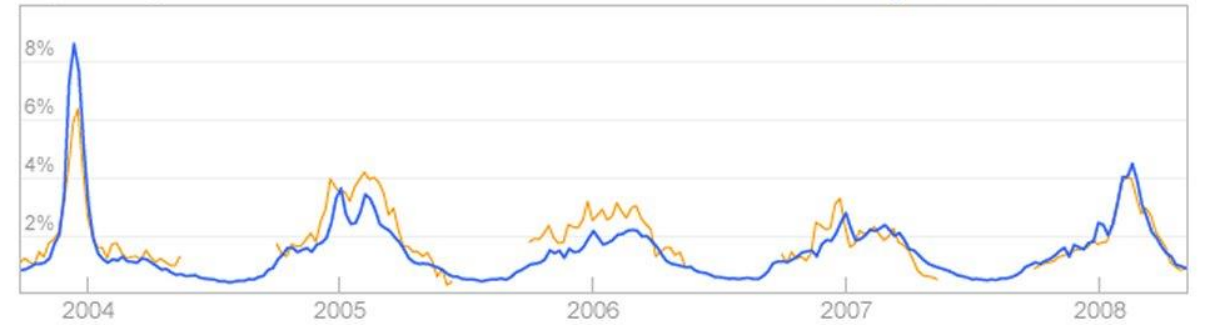
(1) ↑Flu cases (X) is causing ↑Flu searches (Y)



Annual U.S. Flu Activity - Mid-Atlantic Region

ILI percentage

● Google Flu Trends ● CDC Data



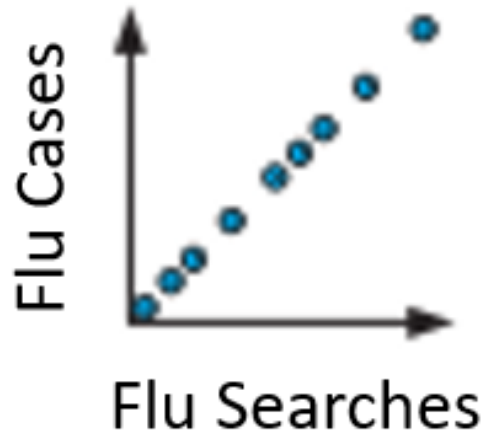
↑ flu cases (X) → ↑ flu searches (Y).

They are **positively** correlated.

This means one of the following is happening:

(1) ↑Flu cases (X) is causing ↑Flu searches (Y)

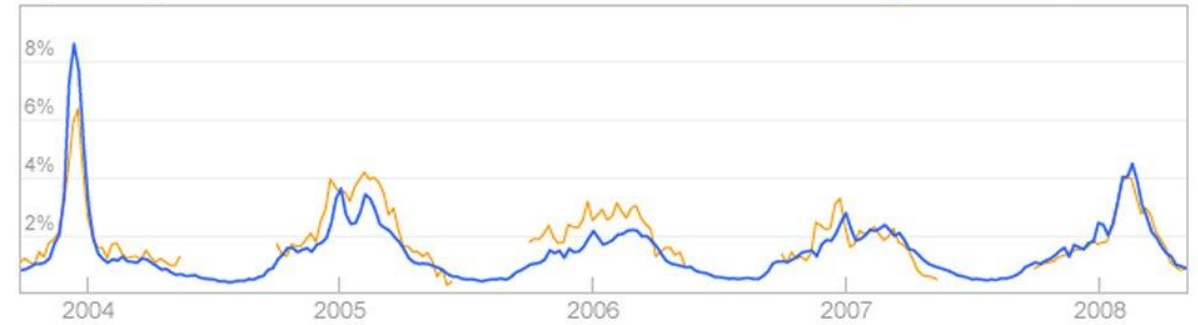
(2) ↑Flu searches (Y) is causing ↑Flu cases (X)



Annual U.S. Flu Activity - Mid-Atlantic Region

ILI percentage

● Google Flu Trends ● CDC Data



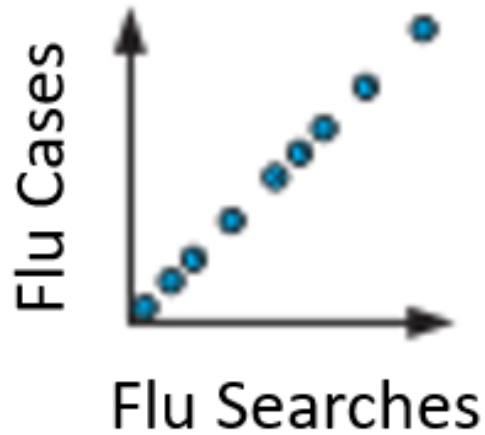
↑ flu cases (X) → ↑ flu searches (Y).

They are **positively** correlated.

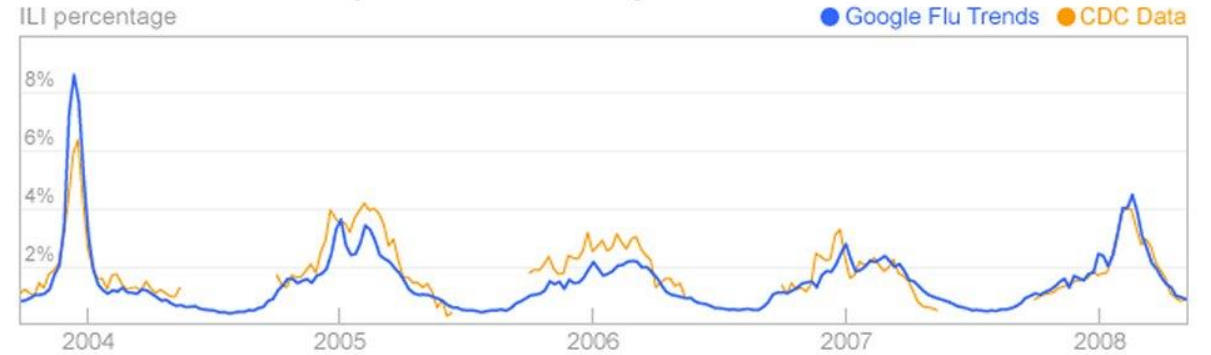
This means one of the following is happening:

- (1) ↑Flu cases (X) is causing ↑Flu searches (Y)
- (2) ↑Flu searches (Y) is causing ↑Flu cases (X)
- (3) A third factor is causing both ↑Flu cases (X) and ↑Flu searches (Y)





Annual U.S. Flu Activity - Mid-Atlantic Region



↑ flu cases (X) → ↑ flu searches (Y).  
They are **positively** correlated.

This means one of the following is happening:

- (1) ↑Flu cases (X) is causing ↑Flu searches (Y)
- (2) ↑Flu searches (Y) is causing ↑Flu cases (X)
- (3) A third factor is causing both ↑Flu cases (X) and ↑Flu searches (Y)

And we don't know which it is without replication, effective sampling and randomization.

## Aluminium in your Tap Water and Alzheimer's Disease link

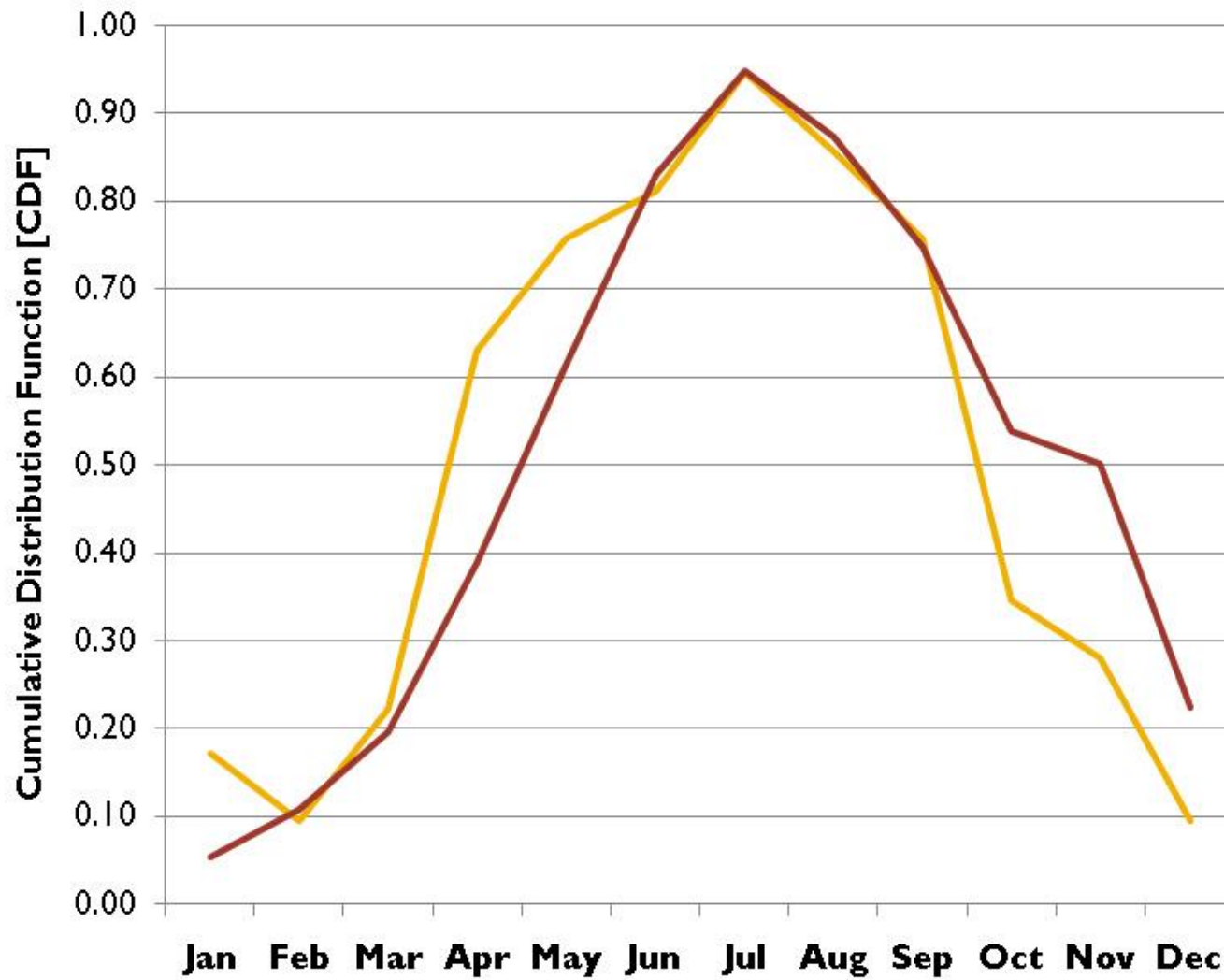


REMOVE IT  
WITH A WATER  
DISTILLER!



Some evidence that it is  
reverse causation:  
Aluminum in brain is  
caused by Alzheimer's.

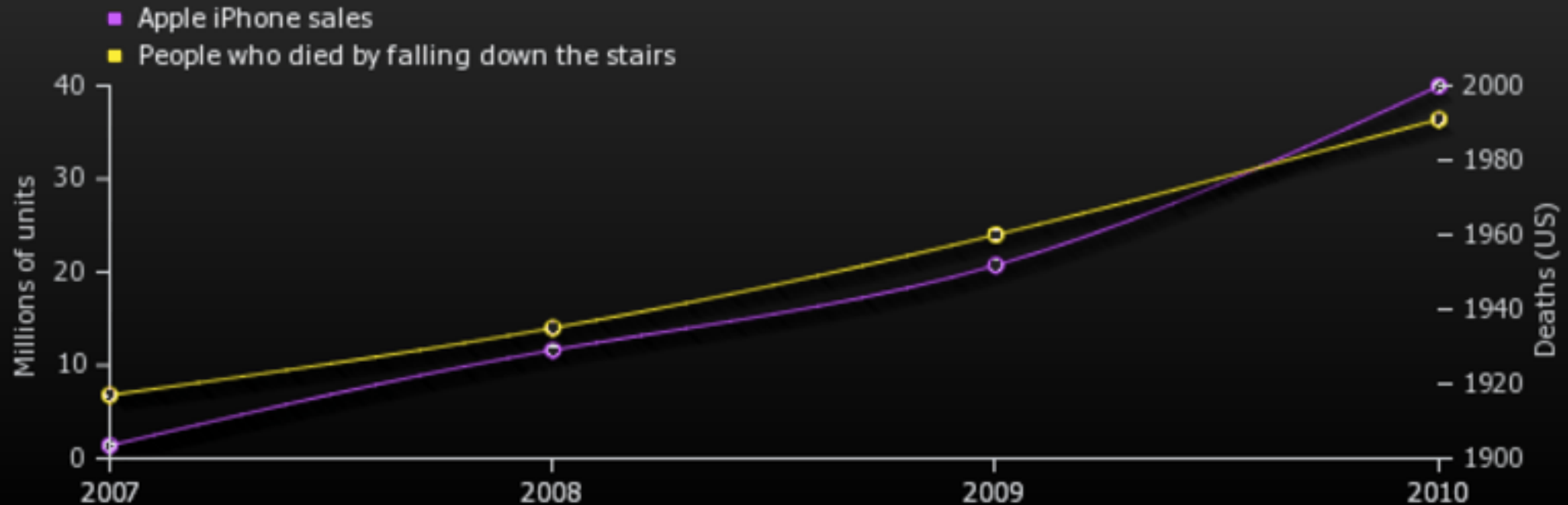
“Fluoride makes your  
body absorb extra  
aluminum. And where  
does the aluminum go?  
Your brain.  
And what metal shows up  
alarmingly in the brains  
of Alzheimer's victims?  
You guessed it.”  
William Douglass, MD,



Spuriousness – third factor is causing both.

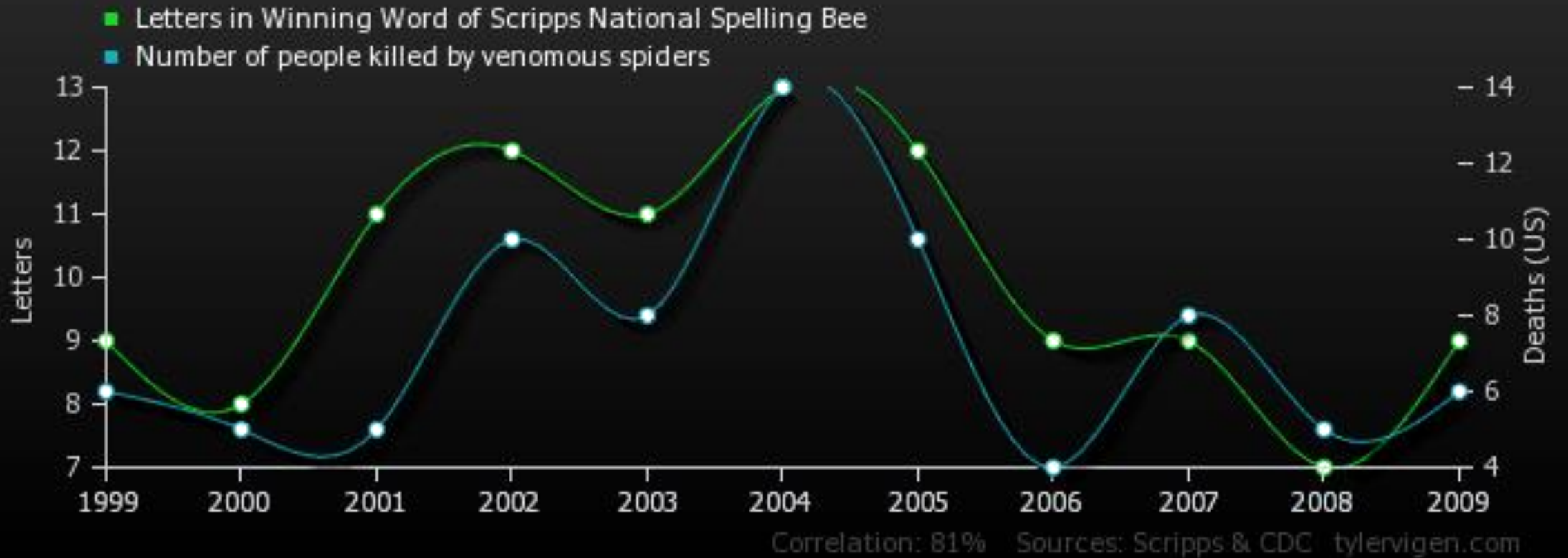
**Violent Crime Index**  
**Ice Cream Sales**

# Apple iPhone Sales Vs People Who Died by Falling down the Stairs



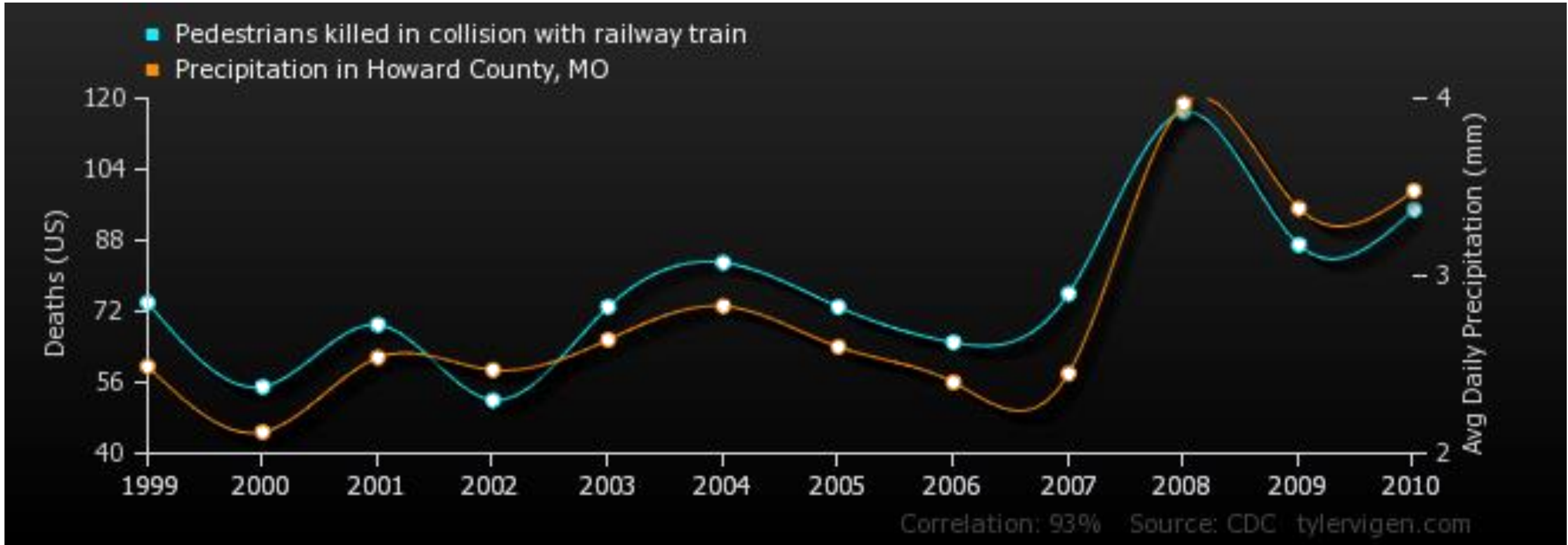
Correlation: 99% Sources: & CDC tylervigen.com

# Spuriousness





# Spuriousness





Another diagram to describe it:



DRY, HOT AND SUNNY  
SUMMER WEATHER



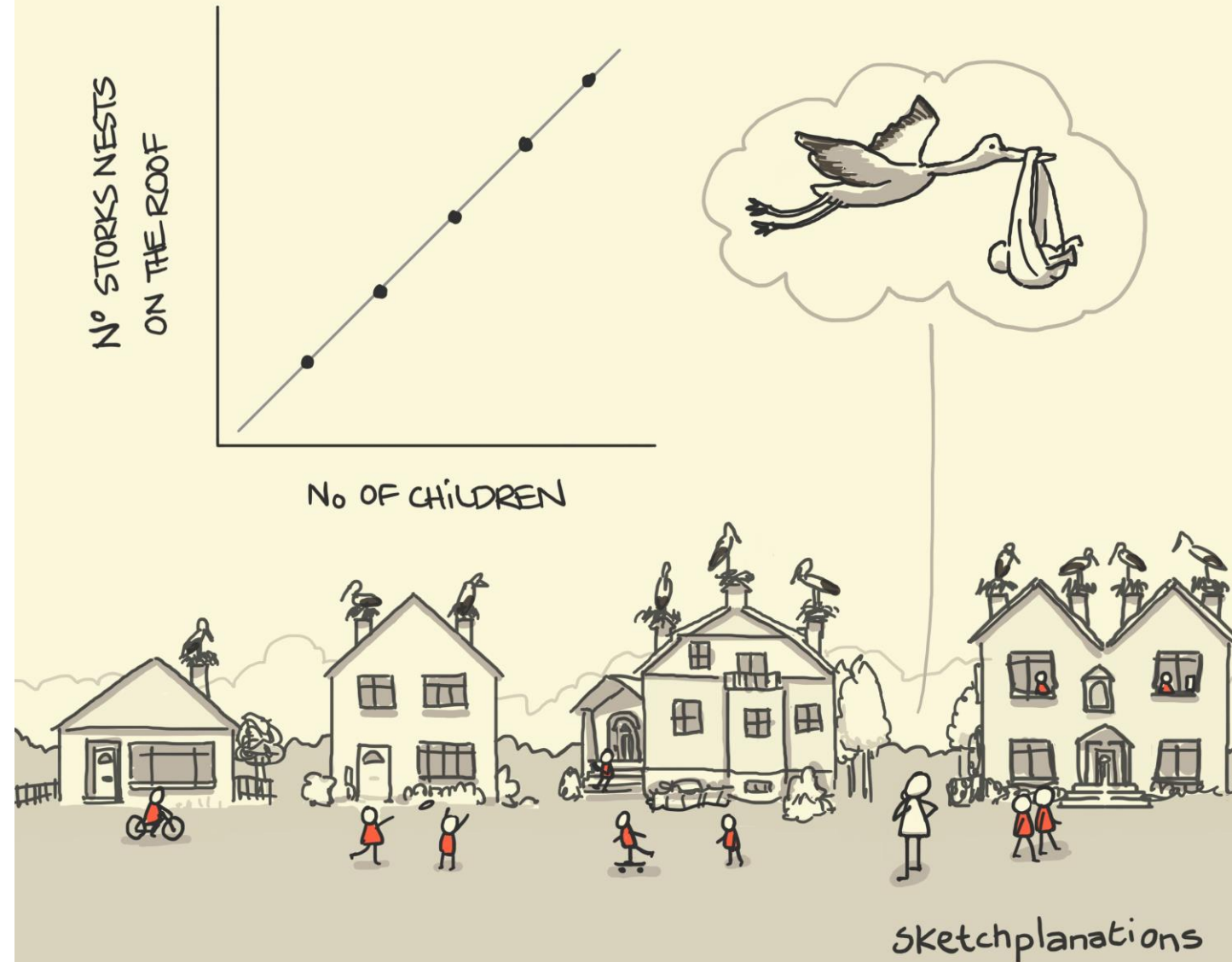
ICE CREAM



SUNBURN

The opening diagram again.

# CORRELATION IS NOT CAUSATION



To prove causation you need:

- A large sample size (replication)
- Random assignment
- Random sampling
- Control groups
- Double blinding
- A logical link between the variables.

## A Frank Statement to Cigarette Smokers

RECENT REPORTS on experiments with mice have given wide publicity to a theory that cigarette smoking is in some way linked with lung cancer in human beings.

Although conducted by doctors of professional standing, these experiments are not regarded as conclusive in the field of cancer research. However, we do not believe that any serious medical research, even though its results are inconclusive should be disregarded or lightly dismissed.

At the same time, we feel it is in the public interest to call attention to the fact that eminent doctors and research scientists have publicly questioned the claimed significance of these experiments.

Distinguished authorities point out:

1. That medical research of recent years indicates many possible causes of lung cancer.
2. That there is no agreement among the authorities regarding what the cause is.
3. That there is no proof that cigarette smoking is one of the causes.
4. That statistics purporting to link cigarette smoking with the disease could apply with equal force to any one of many other aspects of modern life. Indeed the validity of the statistics themselves is questioned by numerous scientists.

We accept an interest in people's health as a basic responsibility, paramount to every other consideration in our business. We believe the products we make are not injurious to health.

We always have and always will cooperate closely with those whose task it is to safeguard the public health.

### TOBACCO INDUSTRY RESEARCH COMMITTEE

5400 EMPIRE

THE AMERICAN TOBACCO COMPANY  
Paul M. Hahn, President

BENSON & EDWARDS  
Joseph F. Cullen, President

BRIGHT BELL & COMPANY  
F. S. R. ...

BROWN & WILLIAMS  
Timothy P. ...

BURLEY AUCTIONEERS  
Albert C. ...

For more than 300 years tobacco has given solace, relaxation, and enjoyment to mankind. At one time or another during those years critics have held it responsible for practically every disease of the human body. One by one these charges have been abandoned for lack of evidence.

Regardless of the record of the past, the fact that cigarette smoking today should even be suspected as a cause of a serious disease is a matter of deep concern to us.

Many people have asked us what we are doing to meet the public's concern aroused by the recent reports. Here is the answer:

1. We are pledging aid and assistance to the research effort into all phases of tobacco use and health. This joint financial aid will of course be in addition to what is already being contributed by individual companies.
2. For this purpose we are establishing a joint industry group consisting initially of the undersigned. This group will be known as TOBACCO INDUSTRY RESEARCH COMMITTEE.
3. In charge of the research activities of the Committee will be a scientist of unimpeachable integrity and national repute. In addition there will be an Advisory Board of scientists disinterested in the cigarette industry. A group of distinguished men from medicine, science, and education will be invited to serve on this Board. These scientists will advise the Committee on its research activities.

This statement is being issued because we believe the people are entitled to know where we stand on this matter and what we intend to do about it.

RECENT REPORTS on experiments with mice have given wide publicity to a theory that cigarette smoking is in some way linked with lung cancer in human beings.

Although conducted by doctors of professional standing, these experiments are not regarded as conclusive in the field of cancer research. However, we do not believe that any serious medical research, even though its results are inconclusive should be disregarded or lightly dismissed.

At the same time, we feel it is in the public interest to call attention to the fact that eminent doctors and research scientists have publicly questioned the claimed significance of these experiments.

Distinguished authorities point out:

1. That medical research of recent years indicates many possible causes of lung cancer.
2. That there is no agreement among the authorities regarding what the cause is.
3. That there is no proof that cigarette smoking is one of the causes.
4. That statistics purporting to link cigarette smoking with the disease could apply with equal force to any one of many other aspects of modern life. Indeed the validity of the statistics themselves is questioned by numerous scientists.

How can you  
tell who is  
telling the  
truth?



# SMOKING and HEALTH

REPORT OF THE ADVISORY COMMITTEE  
TO THE SURGEON GENERAL  
OF THE PUBLIC HEALTH SERVICE

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE  
Public Health Service

In 1964, the Surgeon General, released the landmark report linked smoking cigarettes with dangerous health effects, including lung cancer and heart disease.

Just one year after the report's initial release, Congress passed the Federal Cigarette Labeling and Advertising Act of 1965, which required the first warning labels on cigarette packages.

## What New York Doctors Think of Smoking

By RICHARD BARR

Will the report of the U. S. Public Health Service, showing a link between cigaret smoking and lung cancer, cause New York doctors to give up cigarets? It will and it won't.

Medical men here appear to have been convinced of the harm of heavy smoking even before the Public Health

"The U. S. Public Health Service report won't influence my smoking habits one bit. I haven't seen the report and I'm not going to read it."

"I love cigarets and I have no intention of giving them up. They make me cough and feel lousy. I believe the warnings against cigarets are accurate."

Service report, but quite a few of them say they will go on smoking just the same.

Of course, they'll tell their patients

"I tell my patients it's not healthy to smoke, and that excessive use is harmful. But so is too much food or driving a car in some circumstances."

"I may think others are liable to go on smoking. But I see the cigaret ads rise every year. I'll continue just as I am."

Dr. Eugene Rappaport, Jamaica, Queens.

to ease up on their smoking or cut it out altogether.

Here are the results of a canvass of New York doctors on whether the U. S.

Public Health Service report will influence their own use of tobacco:

Dr. Eric Andraea, 1129 Madison ave., chest disease specialist at St. Luke's Hospital, and assistant physician at Knickerbocker Hospital:

"I'm not a heavy smoker, about half a pack a day. The Federal report won't make me quit."

"But I definitely know that cigaret smoking is connected with cancer. I

Turn to DOCTORS, Page 4

A COMPLETE  
TV MAGAZINE  
YOURS FREE

New York  
**Journal and American**

4 SECTIONS—112 PAGES  
No. 27,506 SUNDAY, JANUARY 12, 1964 SUNDAY 25 CENTS

SUNDAY EDITION

Today's  
Weather:  
Snow, sleet,  
freezing rain

# THE CIGARET REPORT Links Cancer, Other Diseases

See Here, Men, Eight Ashtrays for Five Smokers!



This is panel of 10 and two who headed committee on smoking. Of 10 on panel five smoke. Note eight ashtrays (circled).

University of Minnesota; Dr. Charles Aubrey LeMaistre, University of Texas Southwestern Medical School; William G. Cochran, professor at Harvard; Dr. Louis Freedland, Fessenden, a professor at Harvard; and Dr. Jacob Furth of Columbia University.

Editor's Report: States of Two Unions

By WILLIAM RANDOLPH HEARST JR., Editor-in-Chief, The Hearst Newspapers

PRESIDENT JOHNSON'S State of the Union message this past week was basically a one-to-one message to the people of this great country. It was a forcefully delivered demand upon the Congress for social legislation combining the outlooks of the Franklin D. Roosevelt New Deal and the John F. Kennedy New Frontier.

With its imposing list of progressive goals to be pursued with speed—ranging from civil rights to elimination of "pockets of poverty"—was a proud badge and a cut-back in production of this material for nuclear weapons.

W. R. HEARST JR. This was also interesting with the top priority given the tax cut, so that

## But Statistics Are Disputed

The Dissent The Report

By EMIL J. STEINHEUSER  
Two doctors nationally known in the field of public health yesterday expressed extremely skeptical views on the value of the U.S. Public Health Service report on the relation between cigaret smoking and cancer.

One was Dr. Charles W. Mayo, one of the founders of the famed Mayo Clinic in Rochester, Minn.

CANT LEGISLATE  
The other reached in a telephone poll by this newspaper.

Dr. Henry S. Cross, chairman of the pathology department at Yale University in New Haven.

Dr. Cross said: "I don't think you can legislate against habits any more."

There is evidence that several other diseases are influenced by smoking.

The government, through the Public Health Service, already is studying "remedial measures."

Education, the health benefits by smoking, will be the essence of the campaign.

It is about nine inches high, inches wide and one inch thick. It weighs more than 1 1/2 pounds and has several hundred pages.

WASHINGTON, Jan. 11 (UPI)—The report of the Surgeon General's committee on smoking and health is about the size of an average book.

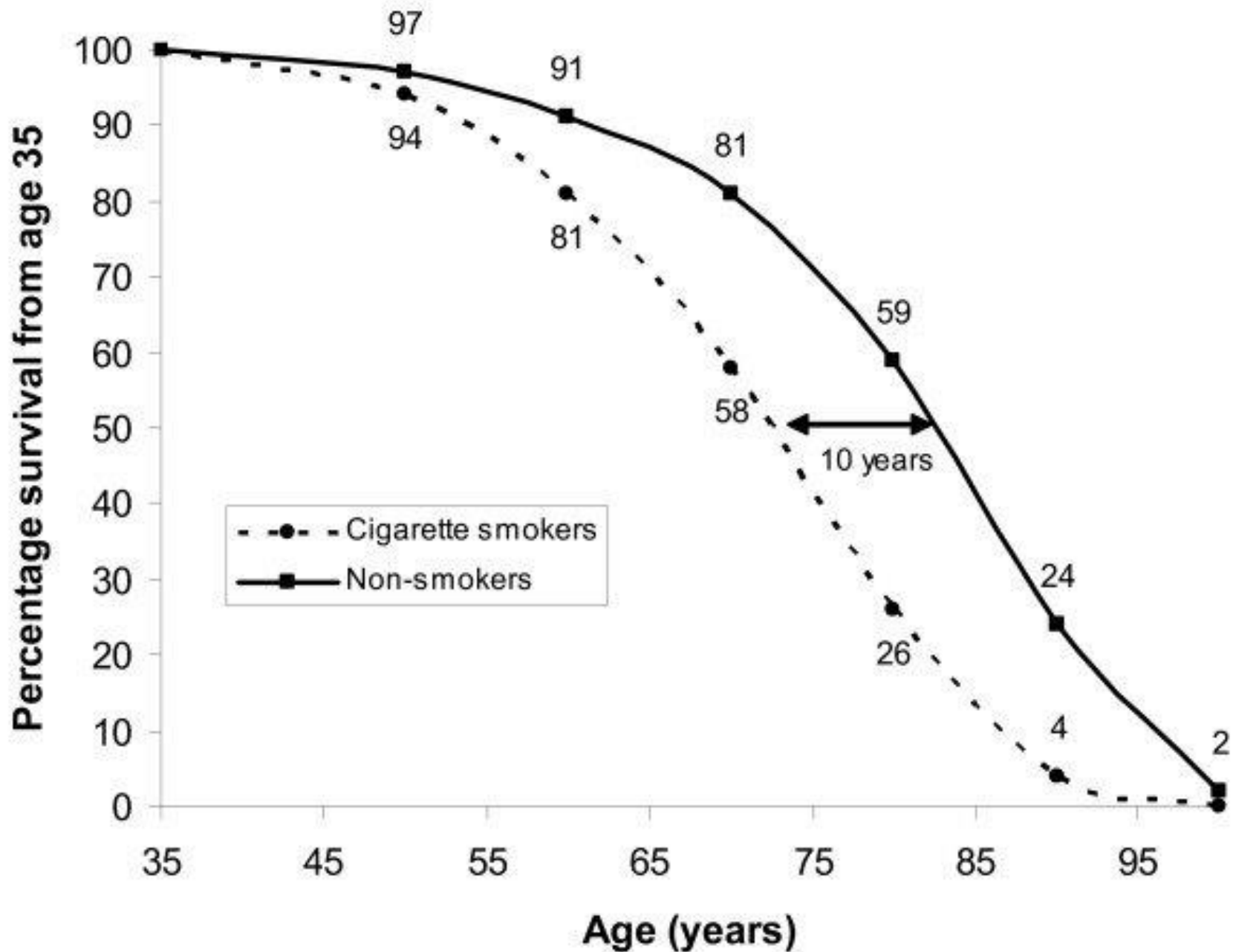
It is about nine inches high, inches wide and one inch thick. It weighs more than 1 1/2 pounds and has several hundred pages.

Today in the Journal-American:  
The Latest Playtime Fashions—Page 25  
The Moment of Decision by Andre March—Page 31

Journal for Living  
Astronaut — 29  
Chas. Barker — 29  
Elinor — 29  
New Yorker — 29  
Ragland — 29  
Sally — 29  
W. J. Sawyer — 29  
Yakovlev — 29

Famous Front Pages—A Fascinating Historic Series Begins Today on Page 17 (Save Them All)





Because of the incredible power of the tobacco lobby, smoking and cancer have some of the strongest causal research that has ever occurred.

Smoking causes cancer.



## Example of Effective Random Assignment

1. Number all pill bottles.
2. (Get a friend to do this) Put in medication or not. Record which numbers have medication and which don't.
3. Get your subject to draw a bottle from the hat. Record which number they used.
4. After all the data is gathered, match up the bottle numbers and find out who had the medication and who didn't.
5. You need a friend to match the bottles with the numbers it so it is "double blinded". The subject doesn't know and you don't either.

BEHAVIORAL ECONOMICS  
RESEARCHERS JAMES HEYMAN AND  
DAN ARIELY HAVE EXAMINED OUR  
SENSITIVITY TO SOCIAL AND  
MARKET NORMS.



WE ASKED PARTICIPANTS TO  
DRAG A CIRCLE INTO A BOX ON A  
COMPUTER SCREEN.





EACH TIME A CIRCLE WAS SUCCESSFULLY  
DRAGGED INTO THE BOX...



...A NEW ONE APPEARED!



FOR THREE MINUTES, DRAG  
AS MANY CIRCLES INTO THE BOX AS  
YOU CAN. ONCE THE TIME IS UP,  
YOU CAN LEAVE.



IT'S AN EASY, IF VERY MONOTONOUS TASK.  
THE NUMBER OF CIRCLES PARTICIPANTS  
DRAGGED REPRESENTED HOW MUCH EFFORT  
THEY WERE WILLING TO PUT INTO THE TASK.





THE PARTICIPANTS  
IN THE STUDY WERE  
RANDOMLY DIVIDED INTO  
THREE DIFFERENT GROUPS.

RESEARCHERS USE  
RANDOMIZED GROUPS SO  
THAT ANY DIFFERENCES  
BETWEEN THE GROUPS ARE  
DUE TO THE EXPERIMENT AND  
NOT DUE TO DIFFERENCES  
BETWEEN THE PARTICIPANTS  
THEMSELVES...



...THIS WAY, ANY INDIVIDUAL  
DIFFERENCES AND QUIRKS  
BALANCE ONE ANOTHER OUT.

(THE LARGER THE  
GROUPS, THE BETTER.)

**RANDOMIZE!**



IN AN EXPERIMENT, EACH OF THESE GROUPS  
IS CALLED A "CONDITION."

OFTEN, ONE OF THE CONDITIONS IS CALLED THE "CONTROL," WHERE NOTHING IS CHANGED AT ALL.



JUST THE  
BASICS OVER  
HERE!

IN THIS STUDY, THE CONDITIONS  
VARIED BY HOW MUCH PARTICIPANTS  
WERE REWARDED FOR THEIR TIME.





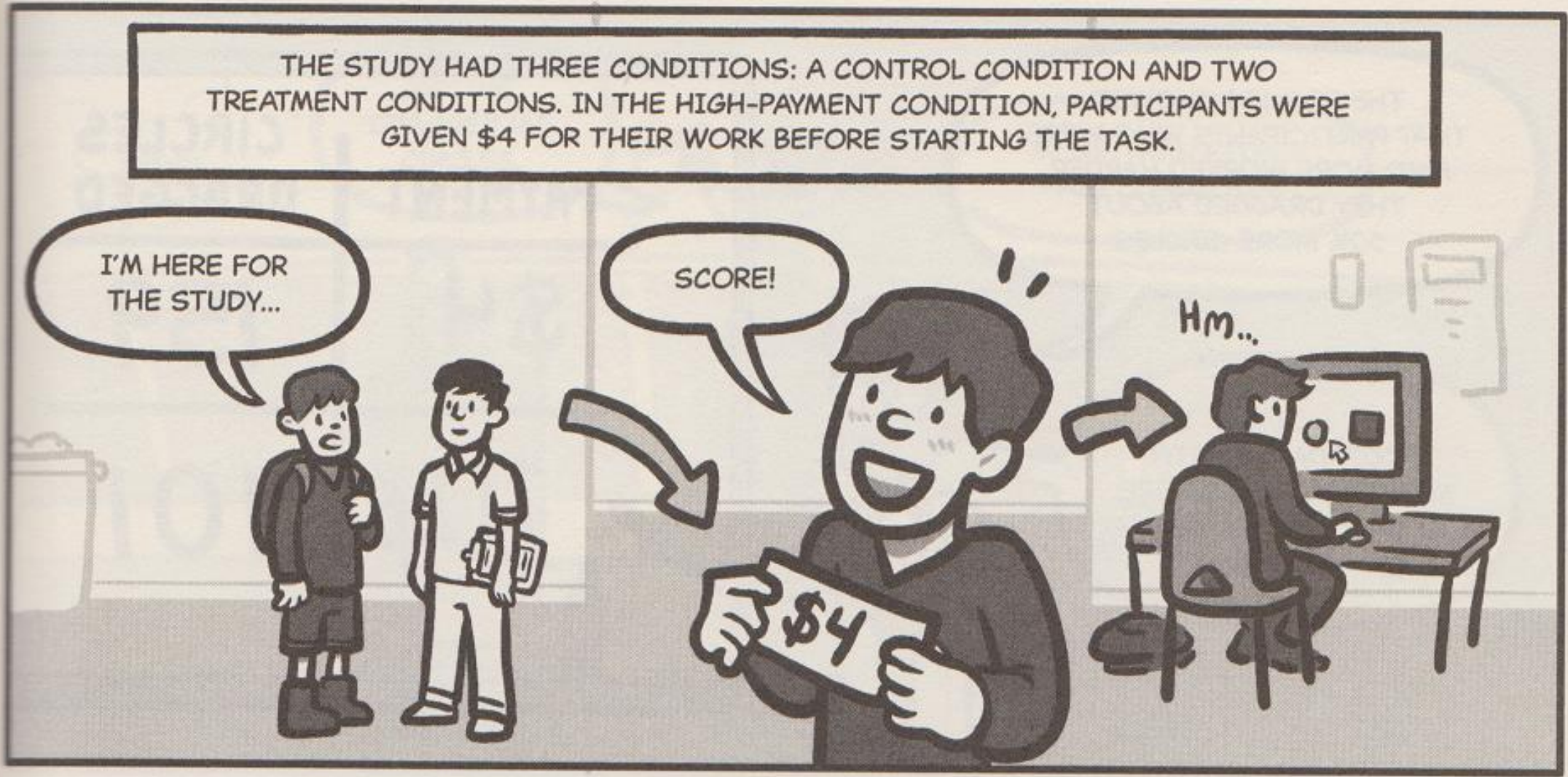
THE STUDY HAD THREE CONDITIONS: A CONTROL CONDITION AND TWO TREATMENT CONDITIONS. IN THE HIGH-PAYMENT CONDITION, PARTICIPANTS WERE GIVEN \$4 FOR THEIR WORK BEFORE STARTING THE TASK.

I'M HERE FOR  
THE STUDY...

SCORE!

Hm...

\$4





PARTICIPANTS IN THE LOW-PAYMENT CONDITION GOT A SMALLER REWARD OF 10¢ BEFORE STARTING THE TASK.

I'M HERE FOR THE STUDY...

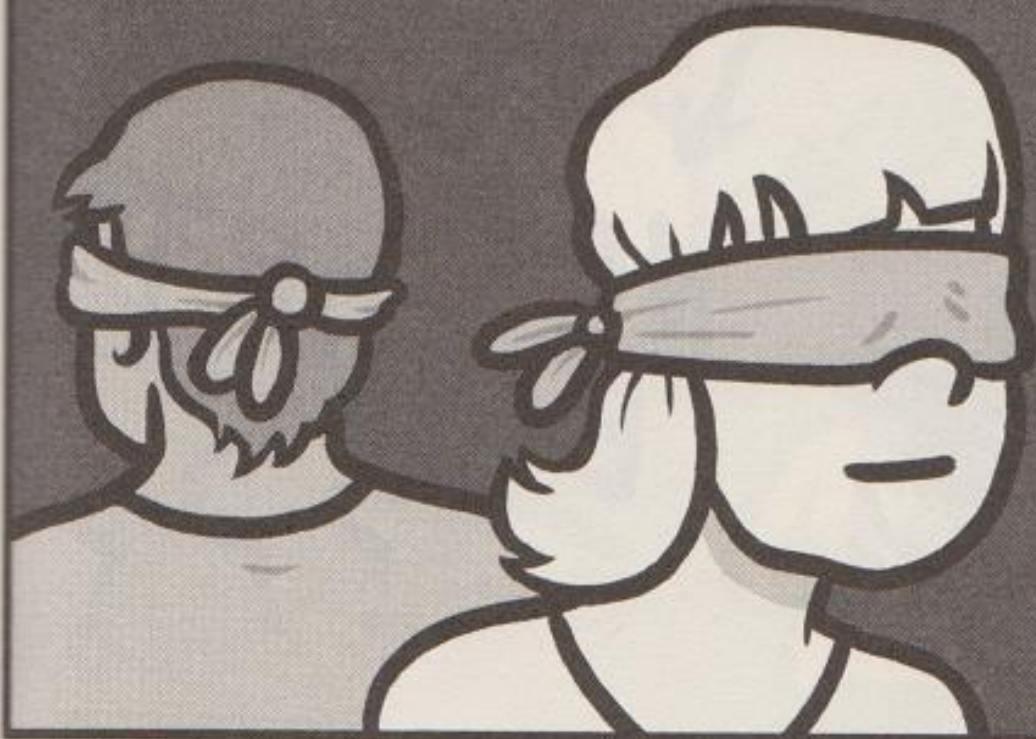
MEH.

Hm...

10



IN MOST EXPERIMENTS, THIS ONE INCLUDED, PARTICIPANTS KNOW ONLY ABOUT THEIR OWN CONDITION - THEY ARE "BLIND" TO THE REST OF THE EXPERIMENT.



SO, NONE OF THE PARTICIPANTS KNEW ANYTHING ABOUT OTHER PARTICIPANTS OR CONDITIONS.






THE RESULTS SHOWED  
THAT PARTICIPANTS WHO WERE  
PAID MORE WORKED HARDER.  
THEY DRAGGED ABOUT  
50% MORE CIRCLES.

THEY BEHAVED IN  
ACCORDANCE WITH  
MARKET NORMS, WHERE  
YOU GIVE PRECISELY WHAT  
YOU ARE PAID FOR.




PAYMENT	CIRCLES DRAGGED
\$4	159
10¢	101





BUT THAT'S NOT ALL!

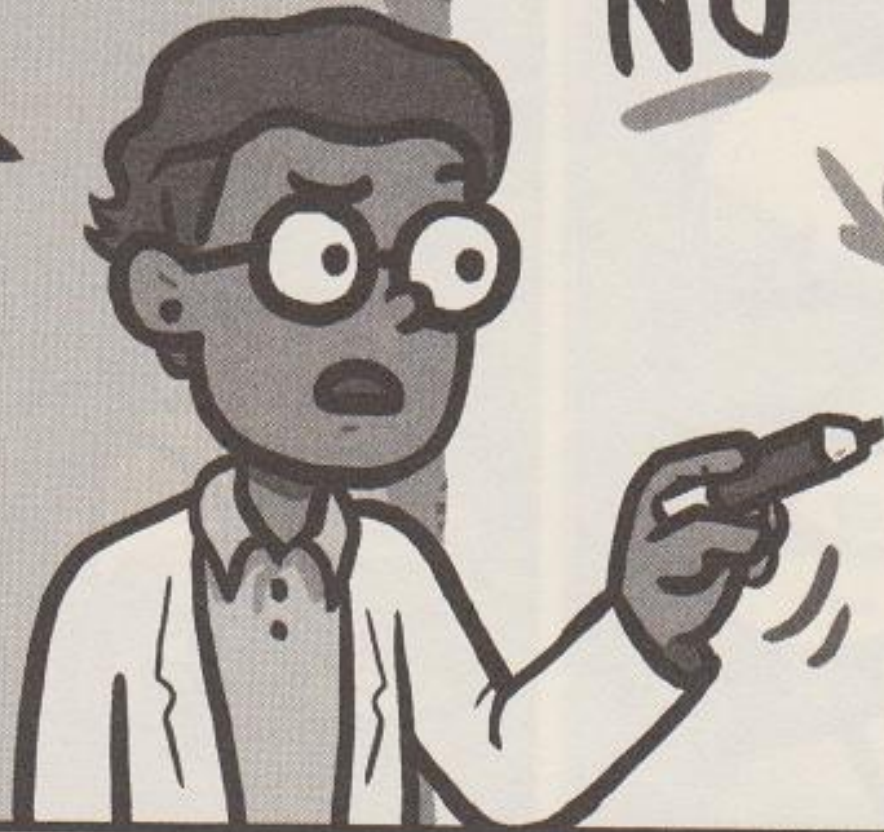
IN THE CONTROL GROUP,  
PARTICIPANTS WEREN'T OFFERED ANY  
COMPENSATION UP FRONT. MONEY WAS  
NEVER MENTIONED AT ALL!



CONSIDER IT  
A FAVOR!



DID THEY, TOO, APPLY  
MARKET NORMS?



NO MONEY  
↓ ↓ ↓  
NO  
EFFORT  
???



NO! WE  
DRAGGED 168  
CIRCLES.

MORE THAN EITHER  
PAID CONDITION!

159  
CIRCLES

168  
CIRCLES

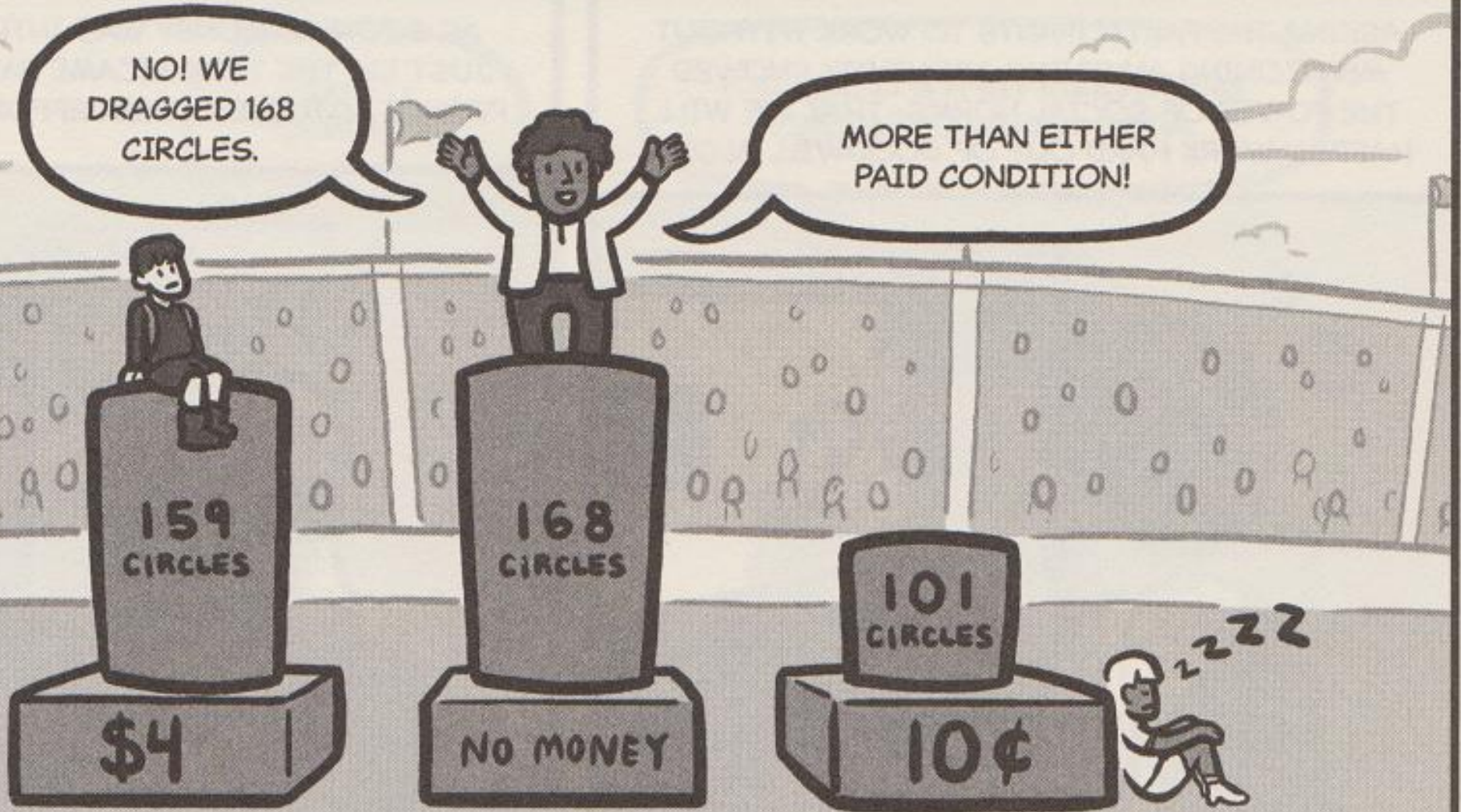
101  
CIRCLES

\$4

NO MONEY

10¢

zzzzz





THIS MEANS THAT PEOPLE  
ACTUALLY WORKED HARDER FOR  
NOTHING THAN FOR SOMETHING!



SOCIAL  
REWARD



HIGHEST  
EFFORT!!



WHETHER AN INTERACTION IS FRAMED IN MARKET OR SOCIAL TERMS STRONGLY SHAPES OUR MOTIVATIONS AND ACTIONS.

I'M SO GLAD I  
COULD HELP THOSE  
SCIENTISTS BY GIVING  
MY TIME!



MY TIME IS  
WORTH MORE THAN 10¢!  
I DON'T FEEL THAT I NEED  
TO WORK VERY MUCH  
FOR JUST 10¢.





ASKING THE PARTICIPANTS TO WORK WITHOUT MENTIONING ANYTHING ABOUT PAY SHOWED THE POWER OF SOCIAL NORMS: THAT WE WILL HAPPILY WORK HARD OUT OF GOODWILL ALONE.



AS SOON AS MONEY WAS INTRODUCED, EVEN JUST 10¢, THE TASK BECAME PAID LABOR - AND PEOPLE ADJUSTED THEIR EFFORT ACCORDINGLY.



# 1. Causal/Descriptive?



1. Causal/Descriptive? – Causal
2. Research Question

1. Causal/Descriptive? – Causal

2. Research Question – Does payment increase motivation?

3. Problem Unit

1. Causal/Descriptive? – Causal
2. Research Question – Does payment increase motivation?
3. Problem Unit – a person
4. Plan Unit –

1. Causal/Descriptive? – Causal
2. Research Question – Does payment increase motivation?
3. Problem Unit – a person
4. Plan Unit – a person in Ariely and Heyman's study from the university's study group.
5. Replication –

1. Causal/Descriptive? – Causal
2. Research Question – Does payment increase motivation?
3. Problem Unit – a person
4. Plan Unit – a person in Ariely and Heyman's study from the university's study group.
5. Replication – unknown
6. Sampling Technique –



1. Causal/Descriptive? – Causal
2. Research Question – Does payment increase motivation?
3. Problem Unit – a person
4. Plan Unit – a person in Ariely and Heyman's study from the university's study group.
5. Replication – unknown
6. Sampling Technique – random from the university's list
7. Diversity Limitations –

1. Causal/Descriptive? – Causal
2. Research Question – Does payment increase motivation?
3. Problem Unit – a person
4. Plan Unit – a person in Ariely and Heyman's study from the university's study group.
5. Replication – unknown
6. Sampling Technique – random from the university's list
7. Diversity Limitations – had to be on university's list, probably university students, in early 20s.
8. Random Assignment –

1. Causal/Descriptive? – Causal
2. Research Question – Does payment increase motivation?
3. Problem Unit – a person
4. Plan Unit – a person in Ariely and Heyman's study from the university's study group.
5. Replication – unknown
6. Sampling Technique – random from the university's list
7. Diversity Limitations – had to be on university's list, probably university students, in early 20s.
8. Random Assignment – to 3 groups: payments of \$4, \$0.05, none
9. Variable –



1. Causal/Descriptive? – Causal
2. Research Question – Does payment increase motivation?
3. Problem Unit – a person
4. Plan Unit – a person in Ariely and Heyman's study from the university's study group.
5. Replication – unknown
6. Sampling Technique – random from the university's list
7. Diversity Limitations – had to be on university's list, probably university students, in early 20s.
8. Random Assignment – to 3 groups: payments of \$4, \$0.05, none
9. Variable – (1) payment, (2) number of circles dragged in 3 minutes
10. Calculation –

1. Causal/Descriptive? – Causal
2. Research Question – Does payment increase motivation?
3. Problem Unit – a person
4. Plan Unit – a person in Ariely and Heyman's study from the university's study group.
5. Replication – unknown
6. Sampling Technique – random from the university's list
7. Diversity Limitations – had to be on university's list, probably university students, in early 20s.
8. Random Assignment – to 3 groups: payments of \$4, \$0.05, none
9. Variable – (1) payment, (2) number of circles dragged in 3 minutes
10. Calculation – average number of circles dragged in each group.

One day some papers catch fire in a wastebasket in the Dean's office. Luckily, a physicist, a chemist, and a statistician happen to be nearby. Naturally, they rush in to help. The physicist whips out a notebook and starts to work on how much energy would have to be removed from the fire in order to stop the combustion. The chemist works on determining which reagent would have to be added to the fire to prevent oxidation.

While they are doing this, the statistician is setting fires to all the other wastebaskets in the adjacent offices. "What are you doing?" the Dean demands.

To which the statistician replies, "To solve a problem of this magnitude, you need a large sample size."



Understand why each isn't  
random, then fix it.

Stand at subway  
entrance at Union  
Station at 7:00 am.

Stand at subway  
entrance at Union  
Station at 7:00 am.

Randomized – put  
times and locations  
(not all subways) in a  
hat. Draw them out.  
Go to a few  
locations.

Go to the cafeteria  
period 2. Ask whoever is  
there your question.



Go to the cafeteria  
period 2. Ask whoever is  
there your question.

Randomized - Put rooms,  
hallways, cafeterias and  
times in a computer. Get  
it to randomly select a  
few locations and times.

Ask whoever  
answers the  
phone.

Ask whoever  
answers the  
phone.

Randomized: ask  
to speak to  
whoever's  
birthday is  
closest?

On-line surveys  
(same problem)



## On-line surveys (same problem)

Randomized: Need a bank of email/house addresses. Get a computer to randomly select the people. Email or mail those people. Bug them to get them to respond.

Walk into forest  
and randomly  
select tree.

Walk into forest  
and randomly  
select tree.

Get a map of the forest.  
Blindfold yourself. Throw  
a dart. Go to that  
location, sample the tree  
closest to it.