Homework Check

8.5 -Regression

Correlation vs. Causation

Correlation

A statistical way to measure the relationship between two sets of data.

Means that both things are observed at the same time.

Causation

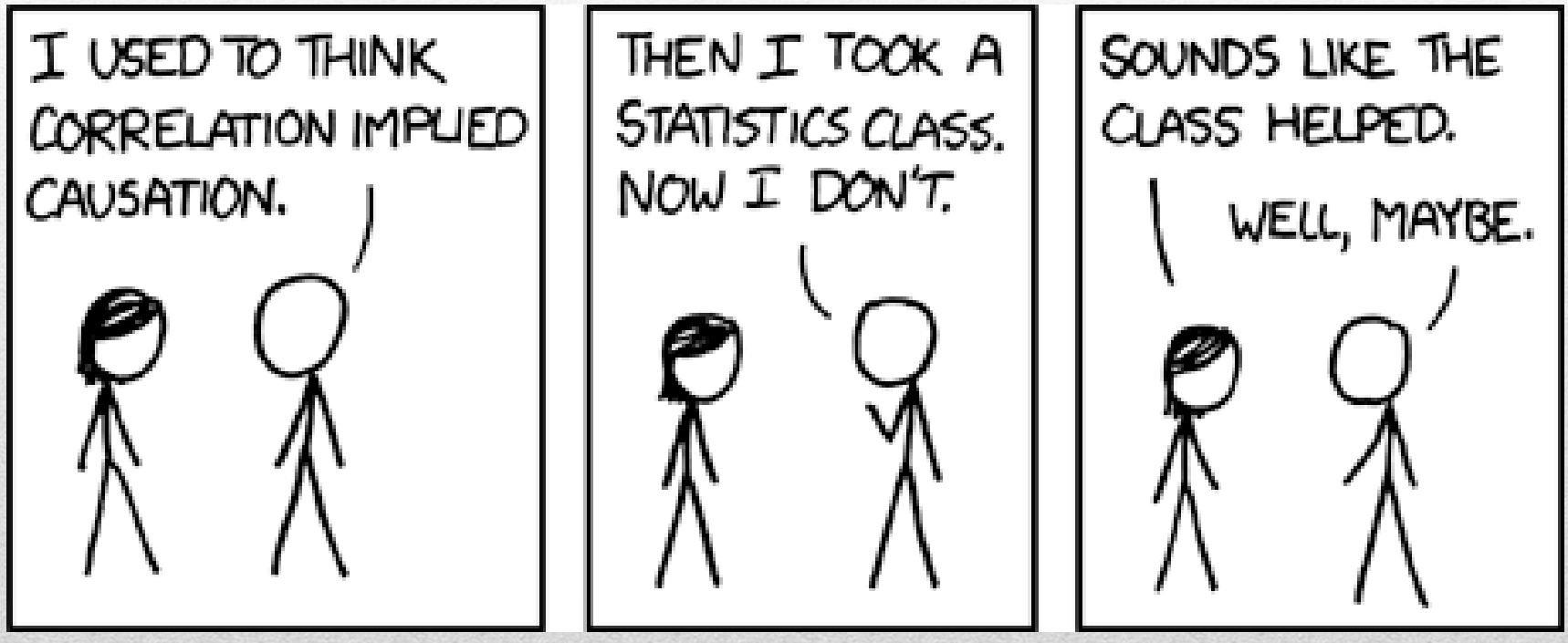
Means that one thing will cause the other.

You can have correlation without causation

There is a correlation (relationship) between the number of firemen fighting a fire and the size of the fire.

(The more firefighters at the scene means that there is a bigger fire.)

However, this doesn't mean that bringing more firemen will cause the size of the fire to increase


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Is it Causation or Correlation?

Ex 1. A recent study showed that college students were more likely to vote than their peers who were not in school.

Correlation

Ex 2. Mrs. Stewart noticed that there was more trash in the hallways after 2nd period than 1st period.

Correlation

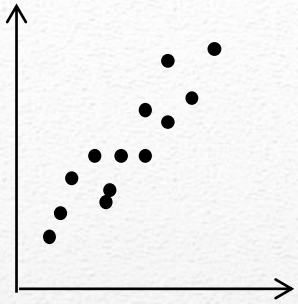
Ex 3. You hit your little sister and she cries

Causation

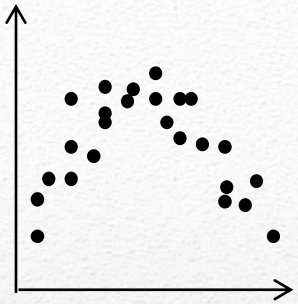
Measuring Correlation

- Correlation is measured by the correlation coefficient, r.
- r is a number between -1 and 1.
- There are 4 traits to correlation:
  1. Form
  2. Direction
  3. Strength
  4. Outliers

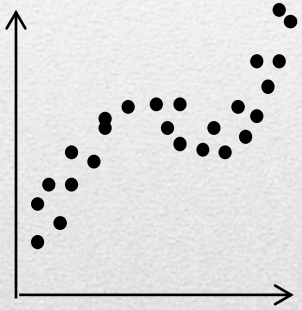
FORM



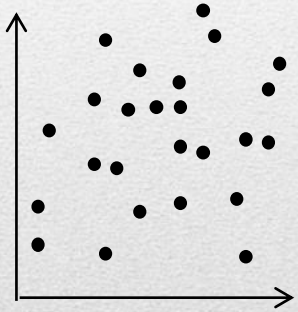
Linear



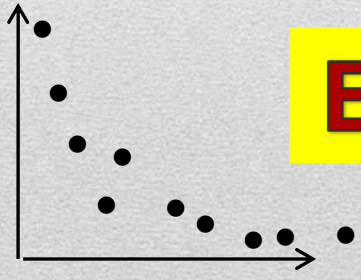
Quadratic



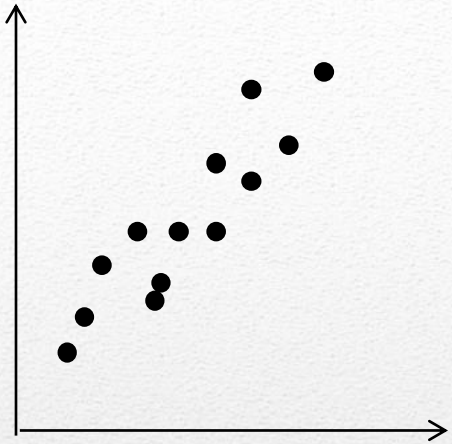
Cubic



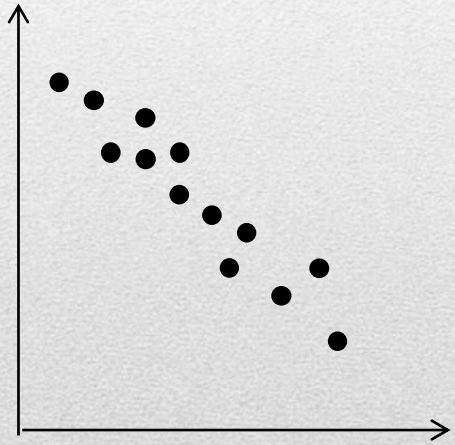
No Correlation



Exponential

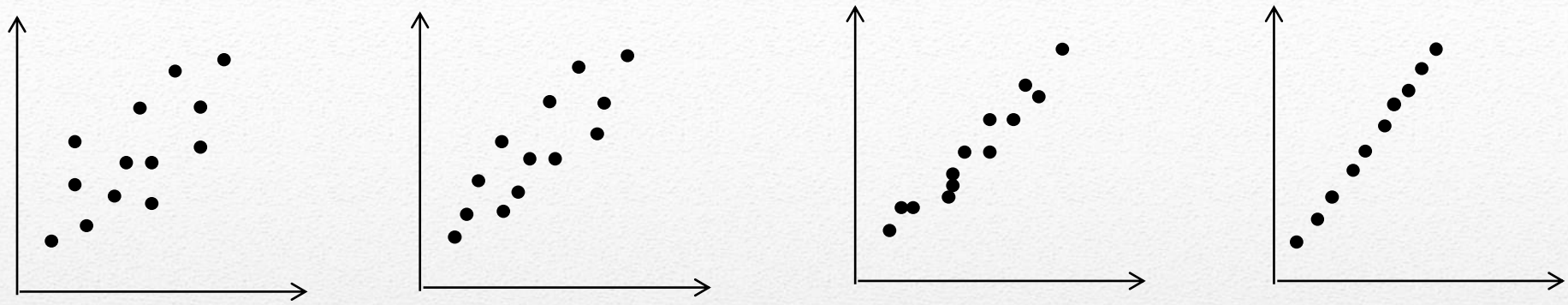


Positive Correlation



Negative Correlation

Direction



Weak ----------> Strong

R value (correlation coefficient)
0 ----------> 1

Strength

Put the correlation coefficients in order from **weakest** to ***strongest***

Ex 1: 0.87, -0.81, 0.43, 0.07, -0.98

0.07, 0.43, -0.81, 0.87, & -0.98

Ex 2: 0.32, -0.65, 0.63, -0.42, 0.04

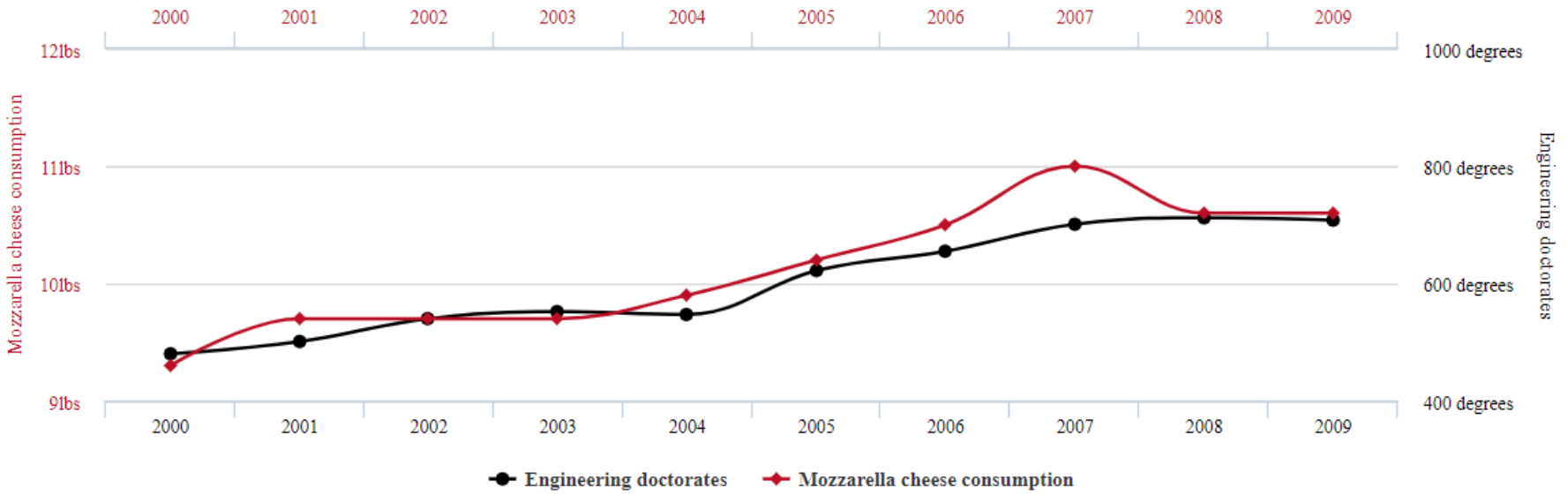
0.04, 0.32, -0.42, 0.63, & -0.65

Per capita consumption of mozzarella cheese

correlates with

Civil engineering doctorates awarded

Correlation: 95.86% (r=0.958648)



tylervigen.com

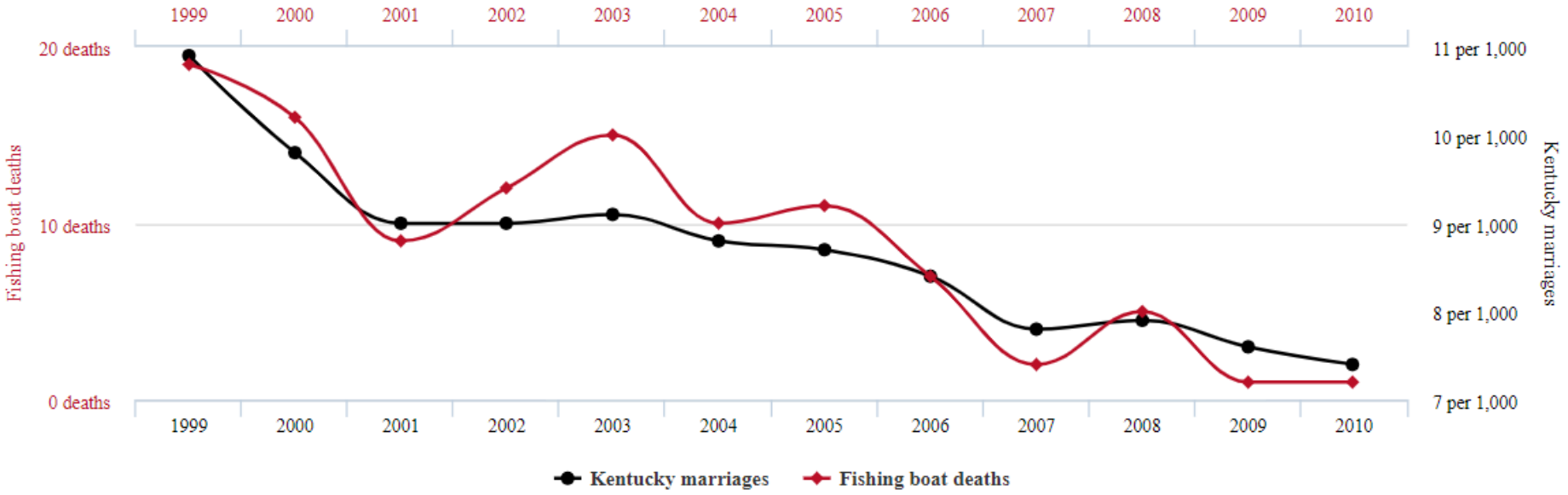
Data sources: U.S. Department of Agriculture and National Science Foundation

People who drowned after falling out of a fishing boat

correlates with

Marriage rate in Kentucky

Correlation: 95.24% (r=0.952407)



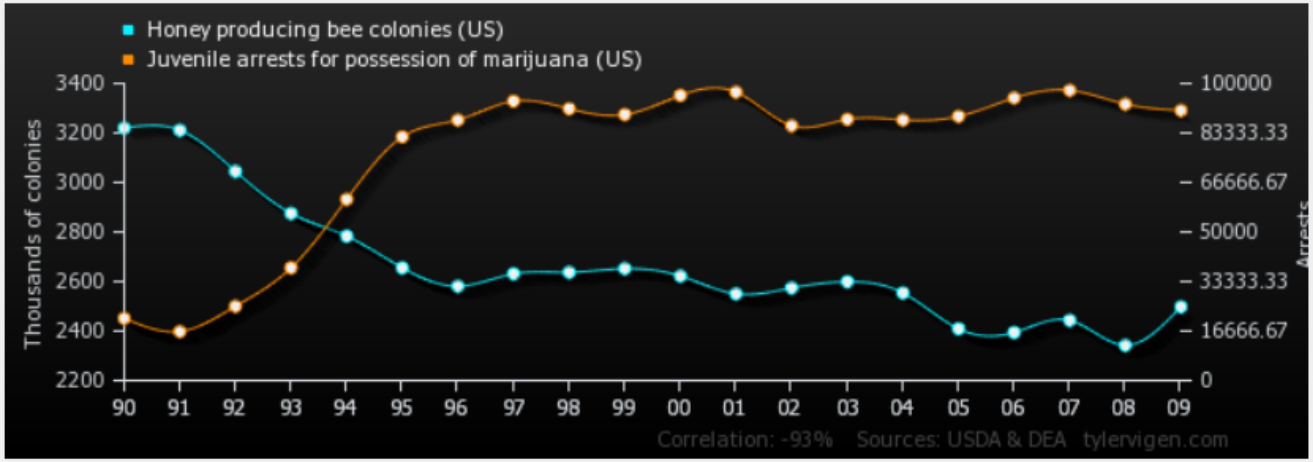
tylervigen.com

Data sources: Centers for Disease Control & Prevention and National Vital Statistics Reports

Honey producing bee colonies (US)

inversely correlates with

Juvenile arrests for possession of marijuana (US)

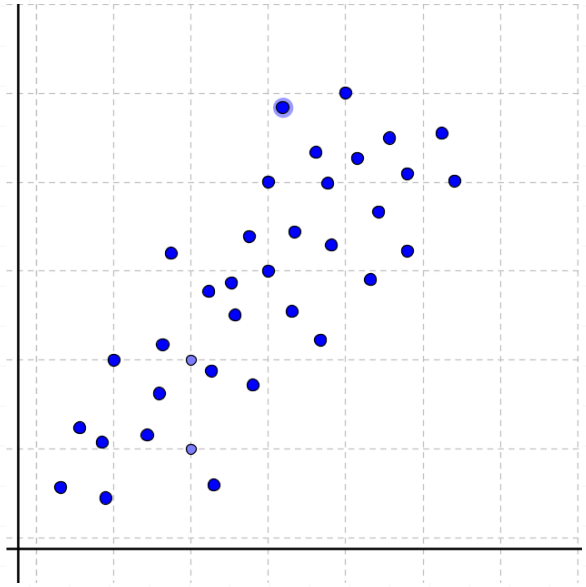


Honey producing bee colonies (US) *Thousands of colonies (USDA)*	'90: 3,220; '91: 3,211; '92: 3,045; '93: 2,875; '94: 2,783; '95: 2,655; '96: 2,581; '97: 2,631; '98: 2,637; '99: 2,652; '00: 2,622; '01: 2,550; '02: 2,574; '03: 2,599; '04: 2,554; '05: 2,409; '06: 2,394; '07: 2,443; '08: 2,342; '09: 2,498
Juvenile arrests for possession of marijuana (US) *Arrests (DEA)*	'90: 20,940; '91: 16,490; '92: 25,004; '93: 37,915; '94: 61,003; '95: 82,015; '96: 87,712; '97: 94,046; '98: 91,467; '99: 89,523; '00: 95,962; '01: 97,088; '02: 85,769; '03: 87,909; '04: 87,717; '05: 88,909; '06: 95,120; '07: 97,671; '08: 93,042; '09: 90,927

Correlation: -0.933389

Match the Correlation Coefficient to the graph

Graph



Correlation Coefficients

-1

-0.5

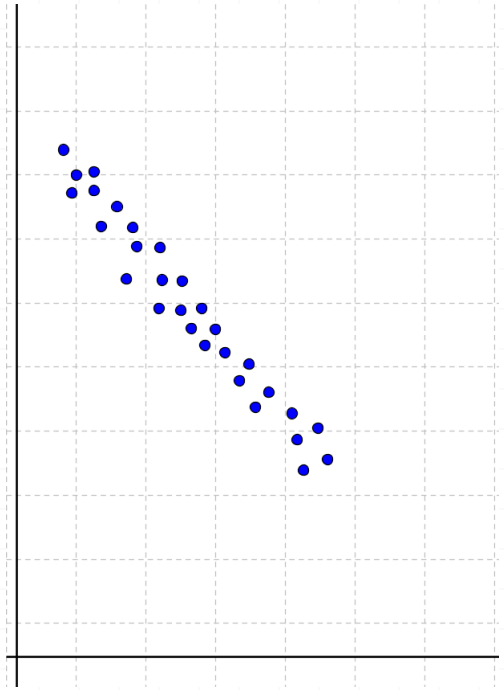
0

0.5

1

Match the Correlation Coefficient to the graph

Graph



Correlation Coefficients

-1

-0.5

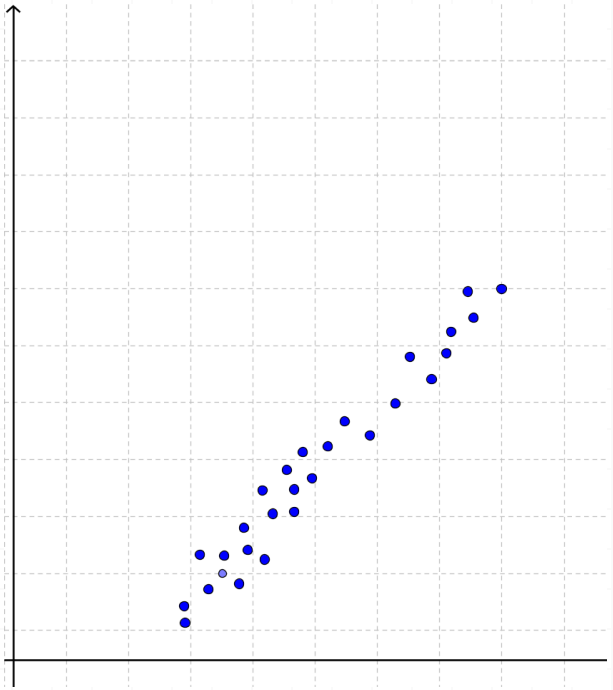
0

0.5

1

Match the Correlation Coefficient to the graph

Graph



Correlation Coefficients

-1

-0.5

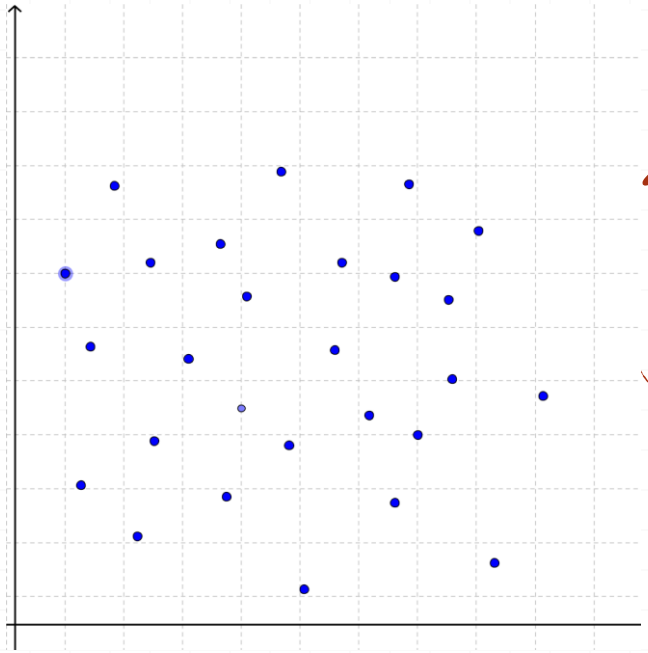
0

0.5

1

Match the Correlation Coefficient to the graph

Graph



Correlation Coefficients

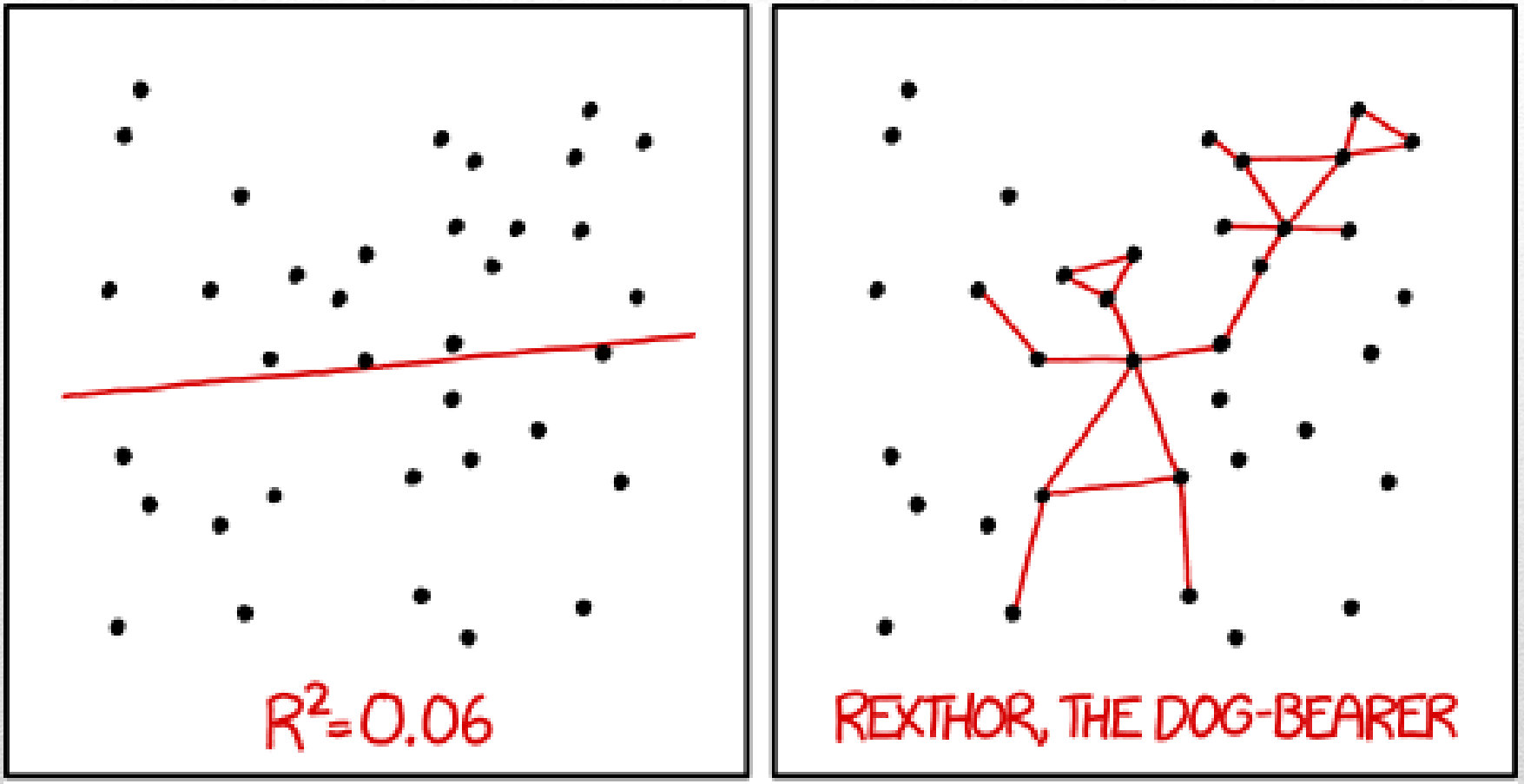
-1

-0.5

0

0.5

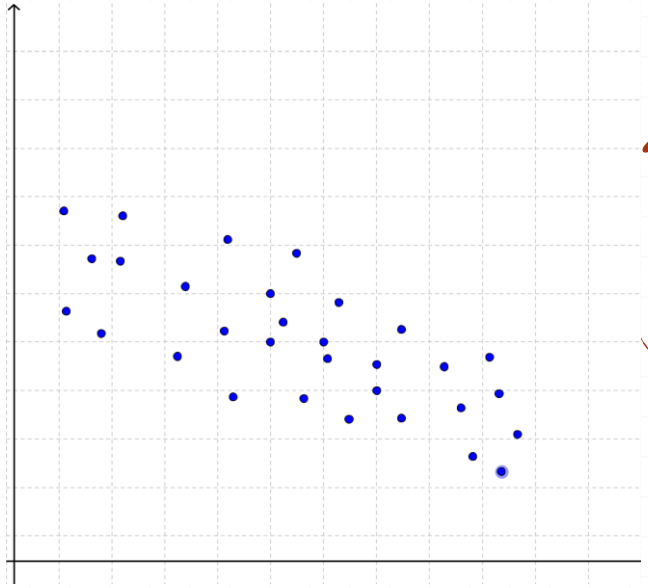
1



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

Match the Correlation Coefficient to the graph

Graph



Correlation Coefficients

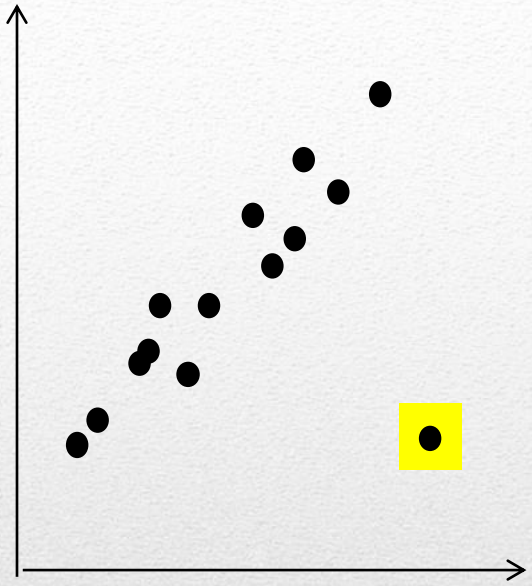
-1

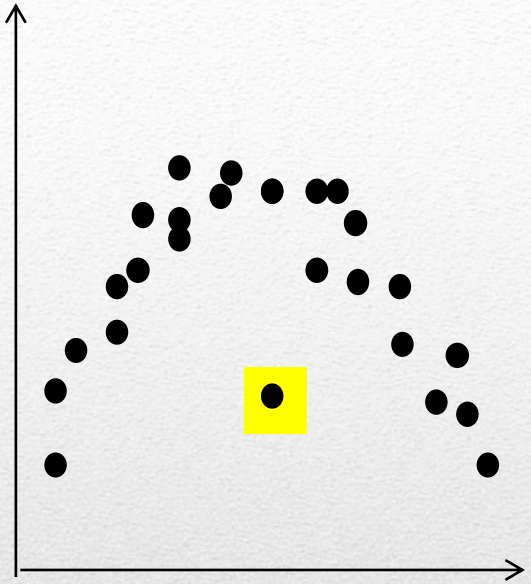
-0.5

0

0.5

1





Data that doesn't fit in

Outliers

Positive, Negative, or No Correlation?

A. The number of hours you work vs. The amount of money in your paycheck

Positive

B. The number of hours workers receive safety training vs. The number of accidents on the job.

Negative

C. The number of students at Hillgrove vs. The number of dogs in Atlanta

No Correlation

Positive, Negative, or No Correlation?

D. The number of heaters sold vs. The months in order from July to February

Positive

E. The number of rice dishes eaten vs. The number of cars on I-75 throughout the day

No Correlation

F. The number of calories burned/lost vs. The amount of hours walked

Positive

Classwork

Correlation Worksheet (with notes)

Homework

Correlation & Causation Worksheet