CS 2810 April 11

Admin:
- HW8 due sunday (HW9 released Tues after)
- Quiz4 scheduling
- TRACE

Content:
Correlation
- what it is
- how its different than correlation

Causation
- its not correlation, even if its easily confused

Conditional Probability
- bayes rule
- independence

Correlation (and covariance): Inuition   (How two values vary together)
The behavior between any two values x and y can be summarized in one of the three ways:

1. as x gets larger y typically gets larger too
   - ex:
     - `x=temp on some day`
     - `y=number of people on the beach on the same day`
   - covariance & correlation is positive
2. as x gets larger y typically doesn't get larger or smaller
   - ex:
     - `x=individual's favorite number`
     - `y=number of hot dogs that individual has eaten in their lifetime`
   - covariance & correlation is zero
3. as x gets larger, y typically gets smaller
   - ex:
     - `x=average speed of driver on 10 mile commute`
     - `y=average commute time of driver on 10 mile commute`
   - covariance & correlation are negative

→ CORR MEASURES JUST THIS
→ COV INCLUDES OTHER FEATURES TOO

So ... whats the difference between correlation and covariance?

How does scaling X impact Sample Cov(X, Y)?

Let:
X be the hours spent doing a HW
Y be the score received on the HW

X	1	2	3
Y	70	80	90

$\bar{X} = 2$

$\bar{Y} = 80$

$$\hat{\sigma}^2_{xy} = \frac{1}{N-1} \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{3-1}\left[(1-2)(70-80) + (2-2)(80-80)\right.$$

$$\left. + (3-2)(90-80)\right]$$

$$= \frac{1}{2}\left[(-1)(-10) + (0)(0) + (1)(10)\right]$$

$$= 10$$

How does scaling X impact Sample Cov(X, Y)?

Let:
X be the minutes spent doing a HW
Y be the score received on the HW

X	60	120	180
Y	70	80	90

$\bar{x} = 120$

$\widehat{y} = 80$

$$\hat{\sigma}^2_{xy} = \frac{1}{N-1} \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{3-1} \left[(60-120)(70-80) + (120-120)(80-80) \right.$$

$$\left. + (180-120)(90-80) \right]$$

$$= \frac{1}{2} \left[(-60)(-10) + (0)(0) + (60)(10) \right]$$

$$= 600$$

How does scaling X impact Sample Cov(X, Y)?

Let c be a scalar while (x_i, y_i) are paired observations.

$$\text{SAMPLE COV OF } cx_i \text{ AND } y_i = \frac{1}{N-1} \sum_i \left(cx_i - c\bar{x} \right)\left(y - \bar{y} \right)$$

$$= c \; \hat{\sigma}_{xy}$$

SCALING (SAMPLES OR OUTCOMES) OF X BY c

WILL SCALE COVARIANCE BY c TOO!

!!Super big problem!!

Changing the units we measure data in will change our covariance scales
- sample cov(hour study, grade) = 10
- sample cov(min study, grade) = 600
- ... we shouldn't interpet the magnitude of covariance as a "strength" of correlation!

Covariance measures three things:
- standard deviation of x
- standard deviation of y
- correlation between x and y (see "intuition" slide above)

We want a "scale invariant" way of measuring correlation

Scale invariant - a statistic which doesn't change when data is scaled

A BOUND ON COVARIANCE

$$\left| \text{Cov}(x, y) \right| = \left| E\left[\left(x - E[x]\right)\left(y - E[y]\right) \right] \right|$$

$$\leq \left| E\left[\left(x - E[x]\right)^2\right]^{1/2} \; E\left[\left(y - E[y]\right)^2\right]^{1/2} \right|$$

$$= \sqrt{\text{Var}(x) \, \text{Var}(y)} = \sigma_x \, \sigma_y$$

COV MAY NOT BE GREATER IN MAGNITUDE
THAN GEOMETRIC MEAN of VAR(x) VAR(y)

CORRELATION (AKA PEARSON'S CORRELATION COEFFICIENT)

CORRELATION $\rightarrow \rho_{xy} = \dfrac{Cov(xy)}{\boxed{\sigma_x \sigma_y}}$

$\rightarrow$ HOW MUCH COV DO X,Y HAVE COMPARED TO WHAT THEY COULD HAVE?

$-1 \le \rho_{xy} \le 1$

$+\sigma_x \sigma_y$ IS MAX $Cov(xy)$
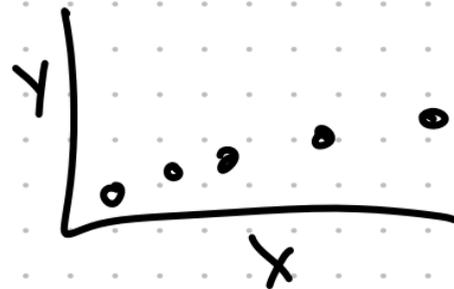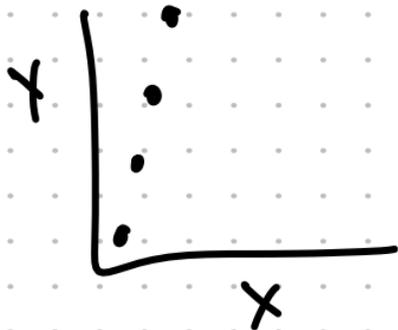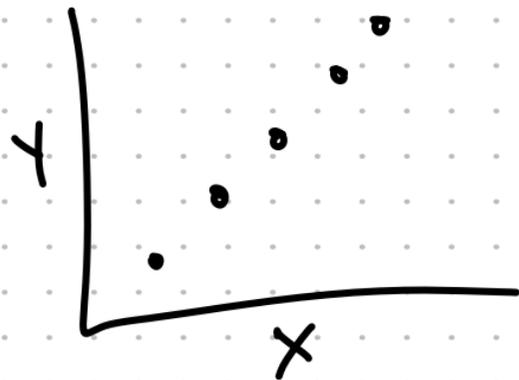
$-\sigma_x \sigma_y$ IS MIN $Cov(xy)$

ESTIMATOR:

$$\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}^2}{\hat{\sigma}_x \hat{\sigma}_y}$$

MAX CORRELATION

$$\text{BOUND } |\text{cov}(xy)| \leq \sigma_x \sigma_y$$

$$1 = \rho_{xy} = \frac{\text{cov}(xy)}{\sigma_x \sigma_y} \implies \text{cov}(xy) = \sigma_x \sigma_y$$

COV IS AS BIG AS POSSIBLE GIVEN SCALE OF X AND Y

All examples have max correlation = 1. This implies that when x increases by some amount then y will always increase by m times that amount. (m is positive)
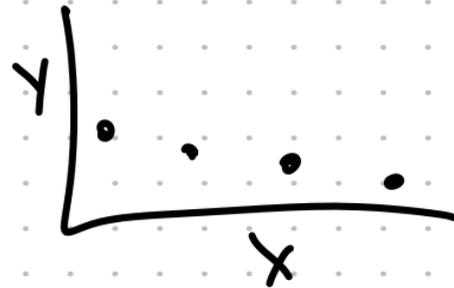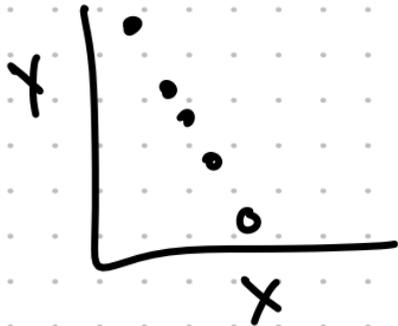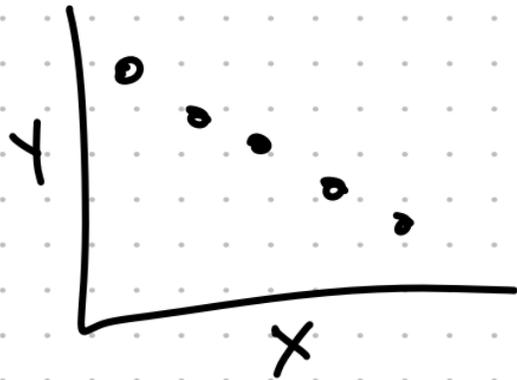
MIN CORRELATION $\qquad$ BOUND $|\text{cov}(xy)| \le \sigma_x \sigma_y$

$$-1 = \rho_{xy} = \frac{\text{cov}(xy)}{\sigma_x \sigma_y} \implies \text{cov}(xy) = -\sigma_x \sigma_y$$

COV IS AS SMALL AS POSSIBLE GWEN SCALE OF X AND Y

All examples have min correlation = 1. This implies that when x increases by some amount then y will always increase by m times that amount (m is negative)

ICA 1

Compute the sample correlation for the four samples below.

Give one sentence which interprets its meaning so a non-technical reader can easily understand.

(Right side of the room, scale x by 2 before computing correlation please!)

$2x$ 14   10   10   6

X	7	5	5	3
Y	2	2	1	3

SAMPLE CORR

SAMPLE COV

$$\hat{\rho}_{xy} = \frac{\overset{\wedge}{\sigma}_{xy}}{\overset{\wedge}{\sigma}_x \overset{\wedge}{\sigma}_y}$$

SAMPLE STD DEVS

ICA 1

$$\hat{\sigma}_Y^2 = \frac{1}{3}\left[(2-2)^2 + (2-2)^2 + (1-2)^2 + (3-2)^2\right] = \frac{2}{3}$$

$$\bar{x} = 5$$
$$\bar{y} = 2$$

$$\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \, \hat{\sigma}_y}$$

$$= \frac{-2/3}{\sqrt{\frac{8}{3}}\sqrt{\frac{2}{3}}}$$

$$= -.5$$

Compute the sample correlation for the four samples below.

Give one sentence which interprets its meaning so a non-technical reader can easily understand.

(Right side of the room, scale x by 2 before computing correlation please!)

X	7	5	5	3
Y	2	2	1	3

$$\hat{\sigma}_{xy}^2 = \frac{1}{N-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{3}\left[(7-5)(2-2) + (5-5)(2-2) + (5-5)(1-2) + (3-5)(3-2)\right]$$

$$= \frac{1}{3}(0+0+0-2) = -2/3$$

$$\hat{\sigma}_x^2 = \frac{1}{N-1}\sum(x_i - \bar{x})^2$$

$$= \frac{1}{3}\left[(7-5)^2 + (5-5)^2 + (5-5)^2 + (3-5)^2\right] = 8/3$$

$$VAR(x) = E\left[\left(X - E[x]\right)^2\right]$$

$$= E[x^2] - E[x]^2$$

What does correlation mean (and what it doesn't) (consistency)

Correlation describes the consistency with which one variable changes in response to another.
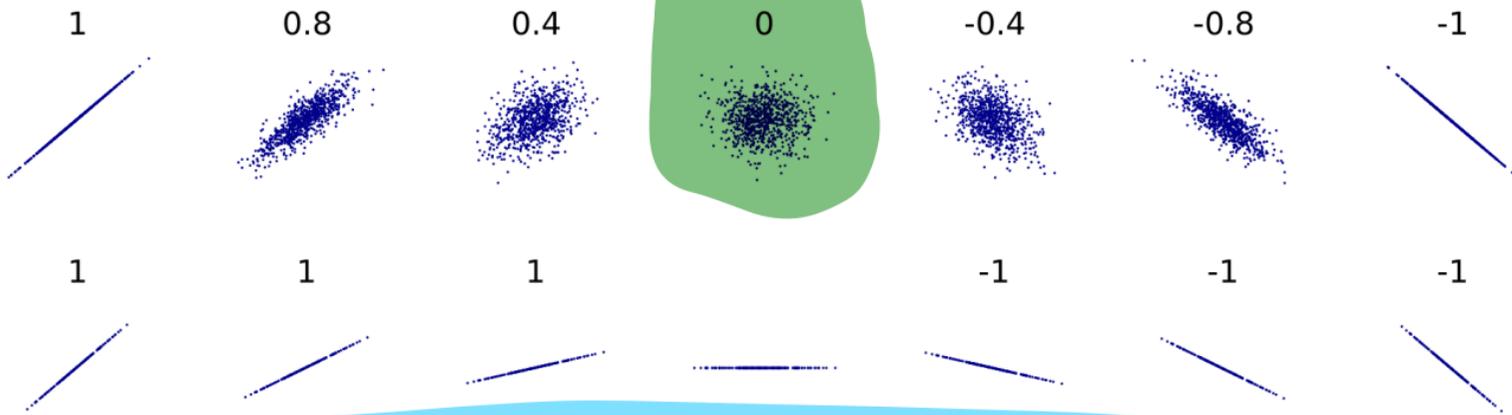The most consistent relationship means whenever one variable goes up by one, the other always changes
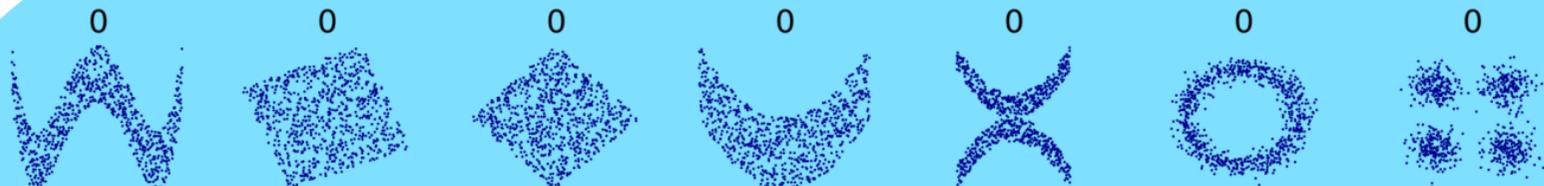by the same amount (forming a line of observations).
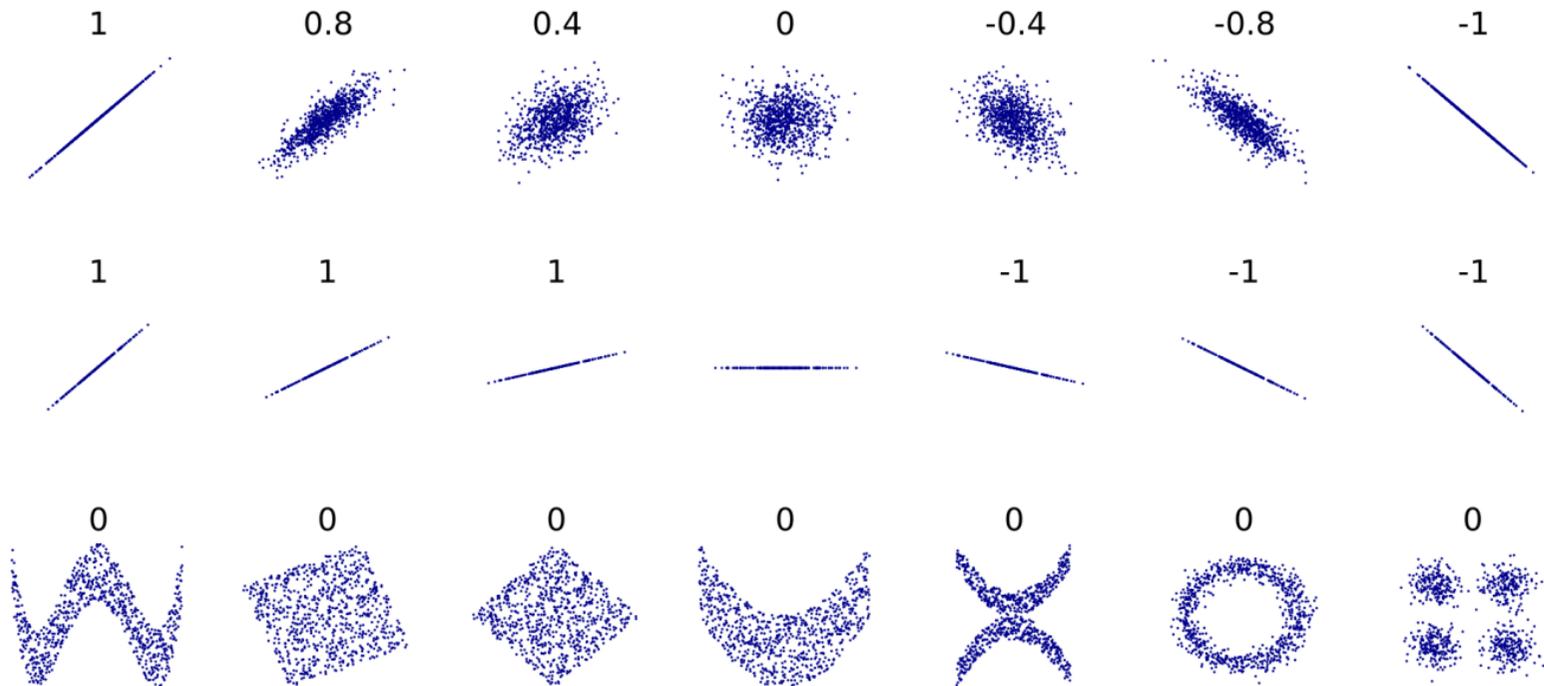
CREDIT
WIKIPEDIA

What does correlation mean (and what it doesn't) (independence)

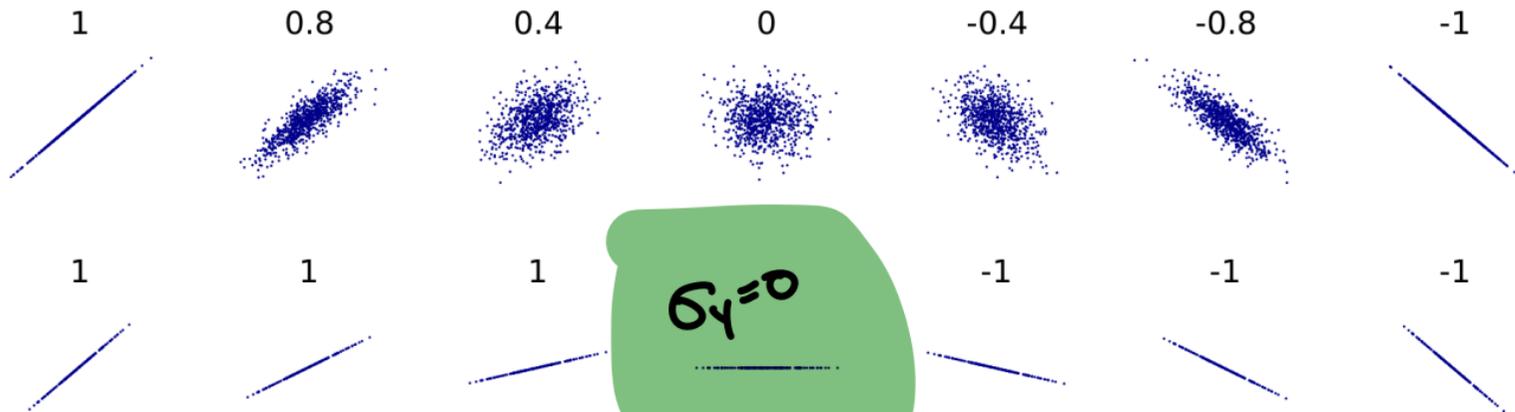If x and y are independent then corr(x, y) = 0

1   0.8   0.4   0   -0.4   -0.8   -1

1   1   1   -1   -1   -1

If corr(x, y) = 0, it may not be true that x and y are independent (see all examples below)
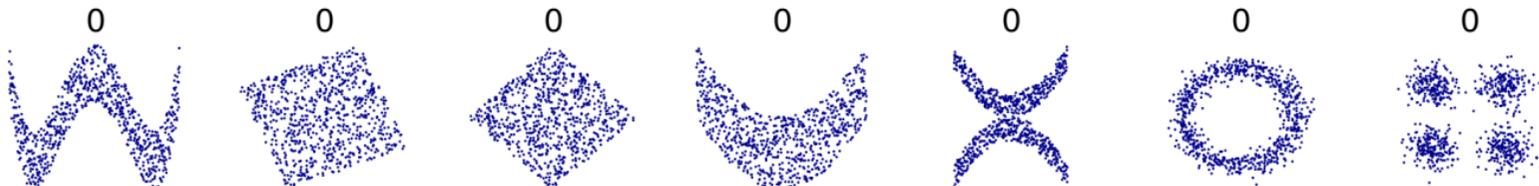
0   0   0   0   0   0   0

What does correlation mean (and what it doesn't) (0 variance case)

Correlation is only defined when each x, y has a positive variance (we'd divide by zero otherwise)

| 1 | 0.8 | 0.4 | 0 | -0.4 | -0.8 | -1 |

$$\rho_{xy} = \frac{cov(xy)}{\sigma_x \sigma_y}$$

| 1 | 1 | 1 | $\sigma_y = 0$ | -1 | -1 | -1 |

If corr(x, y) = 0, it may not be true that x and y are independent (see all examples below)

| 0 | 0 | 0 | 0 | 0 | 0 | 0 |

CLASS PRACTICE

For each pair of random variables below, estimate pearson's correlation coefficient.

x = # hours student studies for final, y = final grade of that student

x = favorite number of student, y = student's best friend's favorite number

x = how many quarters one spends at supermarket, y = weight diff in coin purse @ purchase

x = miles run in the past year, y = average speed of a runner on a 1 mile course

x = internet explorer usage rate, y = covid cases in US (per year ... past 30 years)

In Class Assignment 2

Compute correlation between X_0, X_1

Which feature shows the most consistent relationship with X_0: X_1 or X_2?

$$\rho_{XY} = \frac{COV(XY)}{\sigma_X \sigma_Y}$$

$$VAR(x) = \sigma_x^2$$

$$\Sigma = \begin{bmatrix} 1 & -1.9 & 2 \\ -1.9 & 4 & \leq \\ 2 & \leq & 9 \end{bmatrix} \quad \text{NOT GIVEN}$$

$$\Sigma = \begin{bmatrix} COV(x_0 x_0) & COV(x_0 x_1) & COV(x_0 x_2) \\ COV(x_1 x_0) & COV(x_1 x_1) & COV(x_1 x_2) \\ COV(x_2 x_0) & COV(x_2 x_1) & COV(x_2 x_0) \end{bmatrix}$$

$$\rho_{X_0 X_1} = \frac{COV(x_0 x_1)}{\sigma_{x_0} \sigma_{x_1}} = \frac{-1.9}{\sqrt{1} \cdot \sqrt{4}} = -.95$$

$$\rho_{X_0 X_2} = \frac{COV(x_0 x_2)}{\sigma_{x_0} \sigma_{x_2}} = \frac{2}{\sqrt{1} \sqrt{9}} = \frac{2}{3}$$

Correlation does not imply causality

see silly examples here: https://www.tylervigen.com/spurious-correlations
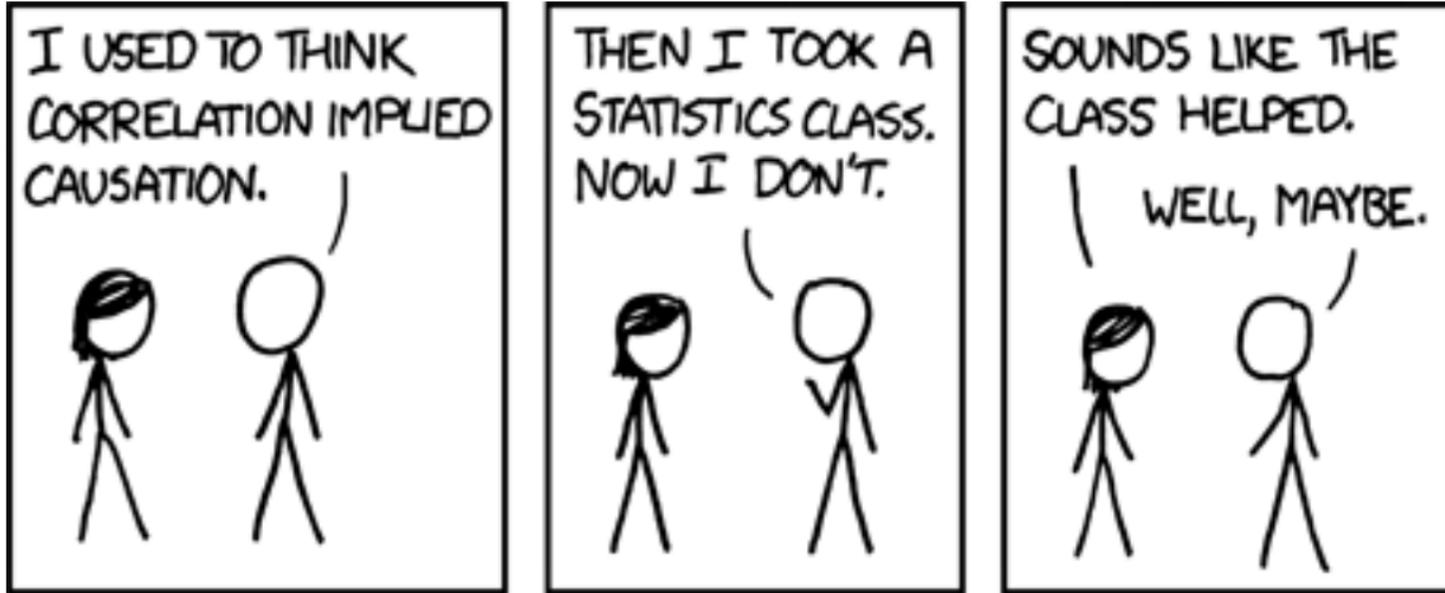
all these correlations are indeed non-zero and show some relationship between variables,
though we shouldn't expect that changing x necessarily impact y
    (discuss titanic hw problem)

Bayes Rule & Conditional Probability

Conditional Probability (intuition / definition)

Whats the probability that a person has covid?

$$P(C=1)$$

Whats the probability that a person has covid given a positive antigen test?

$$P(C=1 \mid A=1)$$

Whats the probability that a person has covid given a negative antigen test?

$$P(C=1 \mid A=0)$$

Whats the probability that an antigen test is negative given a person has covid?

$$P(A=0 \mid C=1)$$

Whats the probability that an antigen test is positive given a person doesn't have covid?

$$P(A=1 \mid P=0)$$

A conditional probability gives the probability of one event given that another has occured.

Let C be a random variable representing whether a person has covid (1=covid, 0=no covid)
Let A be a random variable representing whether that person's antigen test is positive (1=positive, 0=negative)

PROB A HAPPENS
GIVEN B ALREADY
HAS

PROB A,B HAPPEN
TOGETHER

$$P\left(A|B\right) = \frac{P\left(A\,B\right)}{P(B)}$$

PROB B HAPPENS

A = 1   LETTER ON SOCKS
   = 0   OTHERWISE
B = 1   RIGHT SIDE CLASS
   0   OTHERWISE

Estimate each of the probabilities to the right from the observation table below. Give a sentence which explains their meaning so a non-technical reader could easily understand their value.

BITCOIN PRICE MOVEMENT

TWITTER SENTIMENT SCORE (TWEETS w/ BITCOIN HASHTAG)

	B=1 UP	B=-1 DOWN
S=1 ☺	10K	9K
S=0 😐	4K	7K
S=-1 ☹	2K	11K

① $P(S=-1\ B=1) = \dfrac{2}{10+9+4+7+2+11} = 2/43$

② $P(S=-1 \mid B=1) = \dfrac{2}{10+4+2} = \dfrac{1}{8}$

③ $P(B=1 \mid S=-1) = \dfrac{2}{2+11} = \dfrac{2}{13}$

④ $P(B=1) = \dfrac{10+4+2}{10+9+4+7+2+11} = \dfrac{16}{43}$

HOW ARE ① ② ④ RELATED?

Estimate each of the probabilities to the right from the observation table below. Give a sentence which explains their meaning so a non-technical reader could easily understand their value.

$$P\left(S=-1 \mid B=1\right) = \frac{P\left(S=-1 \; B=1\right)}{P\left(B=1\right)}$$

$$\frac{1}{8} = \frac{2/43}{16/43} = \frac{2}{16}$$

① $P\left(S=-1 \; B=1\right) = \frac{2}{10+9+4+7+2+11}$
$$= 2/43$$

② $P\left(S=-1 \mid B=1\right) = \frac{2}{10+4+2} = \frac{1}{8}$

③ $P\left(B=1 \mid S=-1\right) = \frac{2}{2+11} = \frac{2}{13}$

④ $P\left(B=1\right) = \frac{10+4+2}{10+9+4+7+2+11} = \frac{16}{43}$

NOW ARE ① ② ④ RELATED?

Estimate each of the probabilities to the right from the observation table below. Give a sentence which explains their meaning so a non-technical reader could easily understand their value.

BITCOIN PRICE MOVEMENT

TWITTER SENTIMENT SCORE (TWEETS w/ BITCOIN HASHTAG)

	B=1 UP	B=-1 DOWN
S=1 ☺	10K	9K
S=0 😐	4K	7K
S=-1 ☹	2K	11K

② $P\left(S=-1 \mid B=1\right) = \dfrac{2}{10+4+2} = \dfrac{1}{8}$