Marginal Probability Distributions (discrete)

For a discrete joint PDF, there are marginal distributions for each random variable, formed by summing the joint PMF over the other variable.

$$f_X(x) = \sum_y f_{XY}(x,y)$$

$$f_Y(y) = \sum_x f_{XY}(x,y)$$

y = number of times city name is stated	x = number of bars of signal strength			
	1	2	3	$f_Y(y)$ =
1	0.01	0.02	0.25	0.28
2	0.02	0.03	0.20	0.25
3	0.02	0.10	0.05	0.17
4	0.15	0.10	0.05	0.30
$f_X(x)$ =	0.20	0.25	0.55	1.00

Called marginal because they are written in the margins

Figure 5-6 From the prior example, the joint PMF is shown in green while the two marginal PMFs are shown in purple.

Conditional Probability Distributions

Recall that $P(B|A) = \frac{P(A \cap B)}{P(A)}$

$$P(Y{=}y|X{=}x) = P(X{=}x, Y{=}y)/P(X{=}x) =$$
$$= f(x,y)/f_X(x)$$

From Example 5-1

$P(Y{=}1|X{=}3) = 0.25/0.55 = 0.455$

$P(Y{=}2|X{=}3) = 0.20/0.55 = 0.364$

$P(Y{=}3|X{=}3) = 0.05/0.55 = 0.091$

$P(Y{=}4|X{=}3) = 0.05/0.55 = 0.091$

Sum = *1.00*

y = number of times city name is stated	x = number of bars of signal strength			
	1	2	3	$f_Y(y)$ =
1	0.01	0.02	0.25	0.28
2	0.02	0.03	0.20	0.25
3	0.02	0.10	0.05	0.17
4	0.15	0.10	0.05	0.30
$f_X(x)$ =	0.20	0.25	0.55	1.00

Note that there are 12 probabilities conditional on *X*, and 12 more probabilities conditional upon *Y*.

X and Y are Bernoulli variables

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

What is the marginal $P_Y(Y=0)$?

A. 1/6
B. 2/6
C. 3/6
D. 4/6
E. I don't know

Get your i-clickers

X and Y are Bernoulli variables

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

What is the conditional P(X=0|Y=1)?

A. 2/6
B. 1/2
C. 1/6
D. 4/6
E. I don't know

Get your i-clickers

Reminder

Statistically independent events

Always true: $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

- **Two events**

Two events are **independent** if any one of the following equivalent statements is true:

(1) $P(A|B) = P(A)$

(2) $P(B|A) = P(B)$

(3) $P(A \cap B) = P(A)P(B)$

- **Multiple events**

The events $E_1, E_2, \ldots, E_n$ are independent if and only if for any subset of these events $E_{i_1}, E_{i_2}, \ldots, E_{i_k}$,

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = P(E_{i_1}) \times P(E_{i_2}) \times \cdots \times P(E_{i_k})$$

Independence of Random Variables X and Y

- **Random variable independence** means that knowledge of any of the values of *X* does not change probabilities of any of the values of *Y*

- Opposite: **Dependence** implies that some values of *X* influence the probability of some values of *Y*

Independence for Discrete Random Variables

- Remember independence of events (slide 13 lecture 4) : Events are independent if any one of the three conditions are met:
  1) $P(A|B)=P(A \cap B)/P(B)=P(A)$ or
  2) $P(B|A)= P(A \cap B)/P(A)=P(B)$ or
  3) $P(A \cap B)=P(A) \cdot P(B)$

- Random variables independent if **all events** A that $Y=y$ and B that $X=x$ are independent if any one of these conditions is met:
  1) $P(Y=y|X=x)=P(Y=y)$ for any x or
  2) $P(X=x|Y=y)=P(X=x)$ for any y or
  3) $P(X=x, Y=y)=P(X=x)\cdot P(Y=y)$

  for every pair x **and** y

X and Y are Bernoulli variables

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

Are they independent?

A. yes
B. no
C. I don’t know

Get your i-clickers

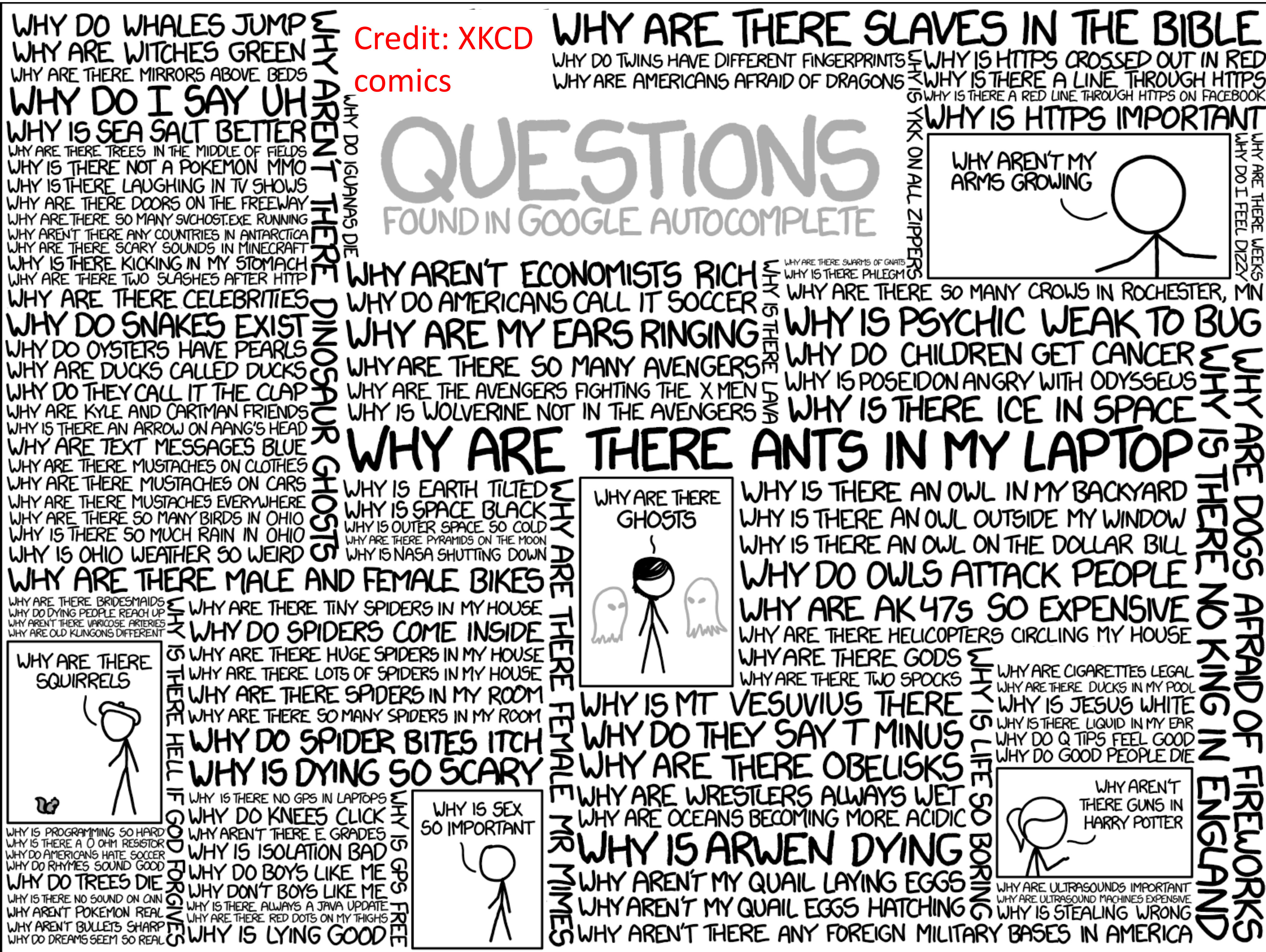
X and Y are Bernoulli variables

	Y=0	Y=1
X=0	1/2	0
X=1	0	1/2

Are they independent?

A. yes
B. no
C. I don't know

Get your i-clickers



Credit: XKCD comics

Joint Probability Density Function Defined

The joint probability density function for the continuous random variables X and Y, denotes as $f_{XY}(x,y)$, satisfies the following properties:

(1) $f_{XY}(x,y) \geq 0$ for all x, y

(2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$

(3) $P\big((X,Y) \subset R\big) = \iint_R f_{XY}(x,y)\,dx\,dy \qquad (5\text{–}2)$

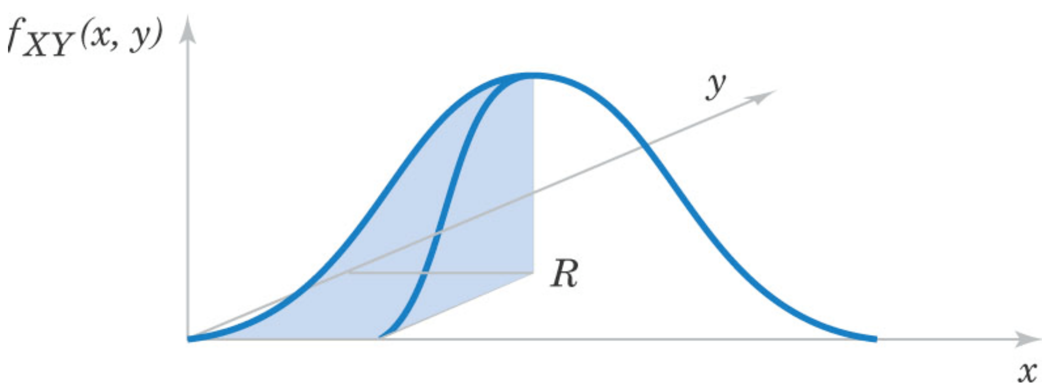


Figure 5-2 Joint probability density function for the random variables X and Y. Probability that (X, Y) is in the region R is determined by the volume of $f_{XY}(x,y)$ over the region R.

Joint Probability Density Function Graph

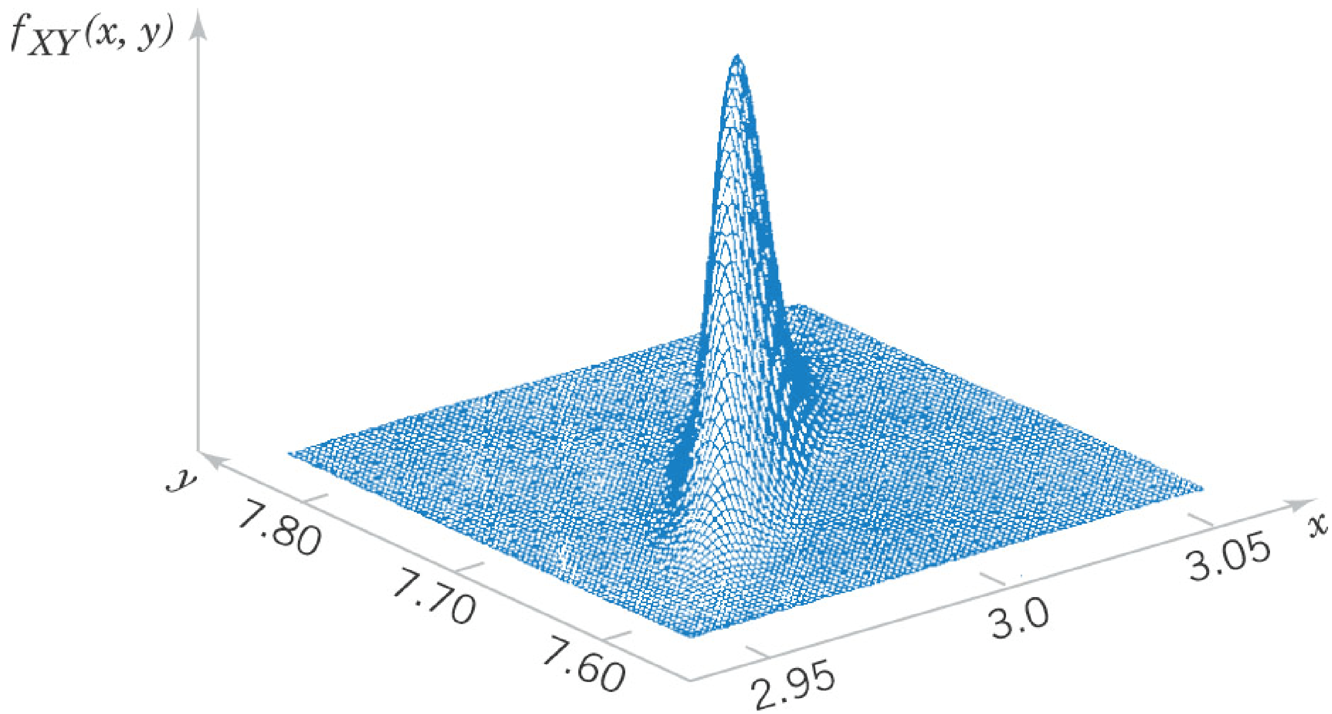


Figure 5-3 Joint probability density function for the continuous random variables *X* and *Y* of expression levels of two different genes. Note the asymmetric, narrow ridge shape of the PDF – indicating that small values in the *X* dimension are more likely to occur when small values in the *Y* dimension occur.

Marginal Probability Distributions (continuous)

- Rather than summing a discrete joint PMF, we integrate a continuous joint PDF.
- The marginal PDFs are used to make probability statements about one variable.
- If the joint probability density function of random variables *X* and *Y* is $f_{XY}(x,y)$, the marginal probability density functions of *X* and *Y* are:

$$f_X(x) = \int_y f_{XY}(x,y)\,dy$$

$$f_Y(y) = \int_x f_{XY}(x,y)\,dx \qquad (5\text{–}3)$$

$$f_X(x) = \sum_y f_{XY}(x,y)$$

$$f_Y(y) = \sum_x f_{XY}(x,y)$$

Conditional Probability Density Function Defined

Given continuous random variables X and Y with
joint probability density function $f_{XY}(x, y)$,
the conditional probability densiy function of Y given X=x is

$$f_{Y|x}(y) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(x,y)}{\int_y f_{XY}(x,y)\,dy} \text{ if } f_X(x) > 0 \qquad (5-4)$$

which satifies the following properties:

(1) $f_{Y|x}(y) \geq 0$

(2) $\int f_{Y|x}(y)dy = 1$

(3) $P(Y \subset B|X = x) = \int_B f_{Y|x}(y)dy$ for any set B in the range of Y

Compare to discrete: *$P(Y=y|X=x)=f_{XY}(x,y)/f_X(x)$*

Conditional Probability Distributions

- Conditional probability distributions can be developed for multiple random variables by extension of the ideas used for two random variables.
- Suppose *p* = 5 and we wish to find the distribution of X_1, X_2 and X_3 conditional on $X_4=x_4$ and $X_5=x_5$.

$$f_{X_1 X_2 X_3|x_4 x_5}(x_1, x_2, x_3) = \frac{f_{X_1 X_2 X_3 X_4 X_5}(x_1, x_2, x_3, x_4, x_5)}{f_{X_4 X_5}(x_4, x_5)}$$

for $f_{X_4 X_5}(x_4, x_5) > 0$.

Independence for Continuous Random Variables

For random variables *X* and *Y*, if any one of the following properties is true, the others are also true. Then *X* and *Y* are independent.

(1) $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$
(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$
(3) $f_{X|y}(y) = f_X(x)$ for all x and y with $f_Y(y) > 0$
(4) $\mathrm{P}(X \subset A, Y \subset B) = P(X \subset A) \cdot P(Y \subset B)$ for any sets A and B in the range of X and Y, respectively. (5−7)

P(Y=y|X=x)=P(Y=y) **for any *x*** or
P(X=x|Y=y)=P(X=x) **for any *y*** or
P(X=x, Y=y)=P(X=x)·P(Y=y) **for any *x* and *y***

Covariation, Correlations

Quick and dirty check for **linear** (in)dependence between variables

Covariance Defined

Covariance is a number quantifying the average *linear* dependence between two random variables.

The covariance between the random variables X and Y, denoted as $\text{cov}(X, Y)$ or σ_{XY} is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y$$

Montgomery, Runger 5th edition Eq. (5−14)

The units of σ_{XY} are the units of X times the units of Y.

Unlike the range of the variance, covariance can be negative: $-\infty < \sigma_{XY} < \infty$.

Covariance – 1 number to measure dependance between random variables

$Cov(X,Y)$ or σ_{XY}

$$\sigma_{XY} = E\left[(X-\mu_X)\cdot(Y-\mu_Y)\right] =$$

$$= E(X\cdot Y) - \mu_X \cdot \mu_Y$$

- $Var(X) = Cov(X,X)$
- If X & Y are independent

$$Cov(X,Y) = E[X-\mu_X]\cdot E[Y-\mu_Y] = 0$$

- $-\infty < Cov(X,Y) < +\infty$

Can be negative!

Covariance and PMF tables

y = number of times city name is stated	x = number of bars of signal strength		
	1	2	3
1	0.01	0.02	0.25
2	0.02	0.03	0.20
3	0.02	0.10	0.05
4	0.15	0.10	0.05

The probability distribution of Example 5-1 is shown.

By inspection, note that the larger probabilities occur as *X* and *Y* move in opposite directions. This indicates a negative covariance.

Covariance and Scatter Patterns

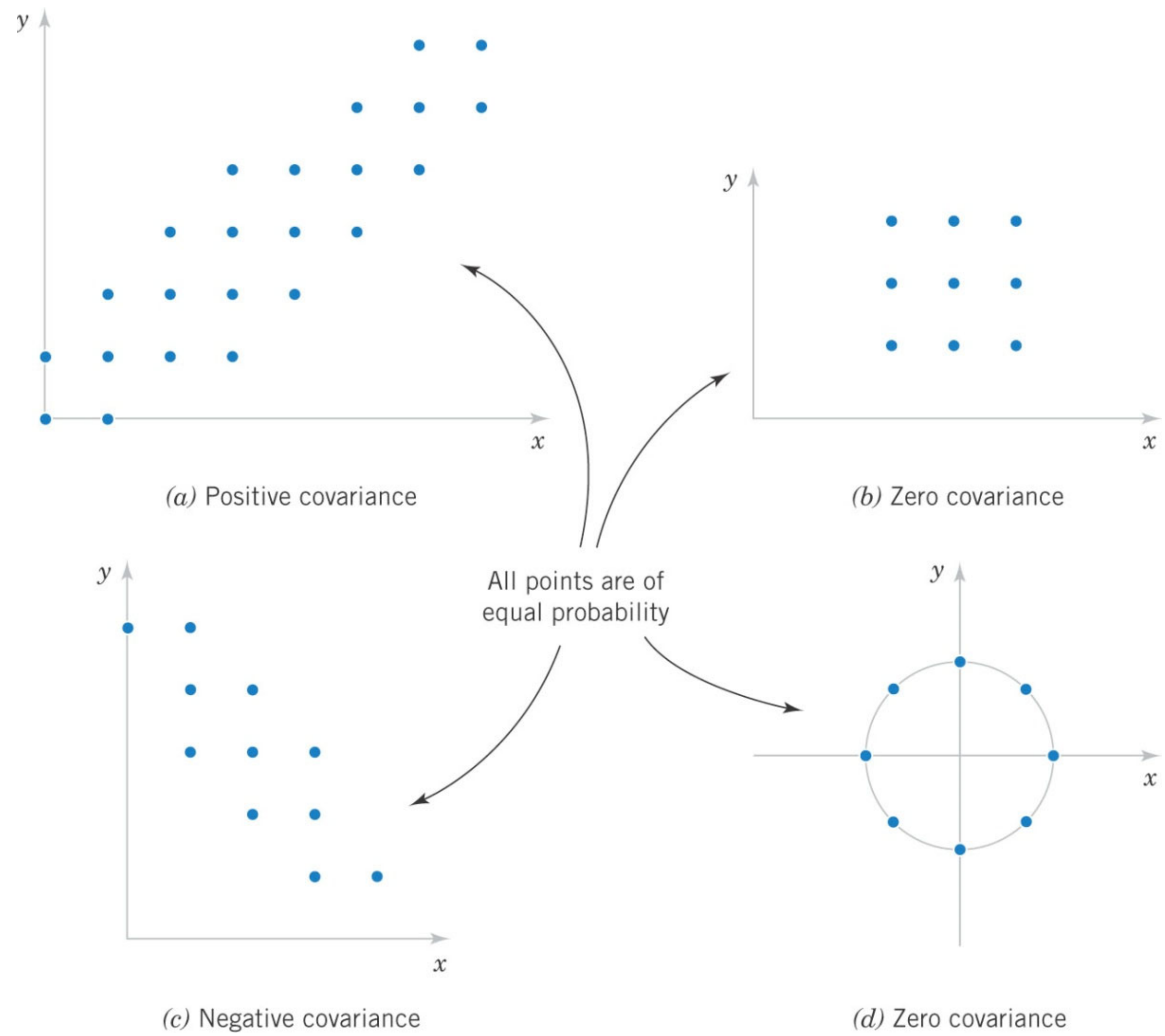


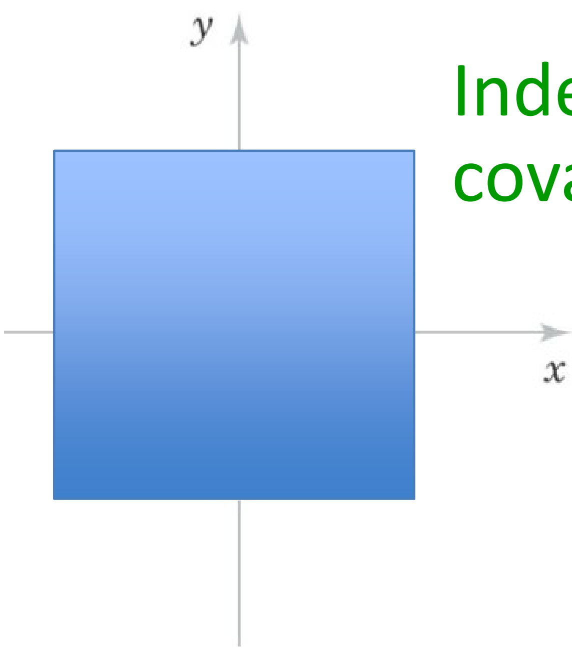
Figure 5-13 Joint probability distributions and the sign of cov(X, Y). Note that covariance is a measure of linear relationship. Variables with non-zero covariance are correlated.

Independence Implies σ=ρ = 0 but <u>not vice versa</u>

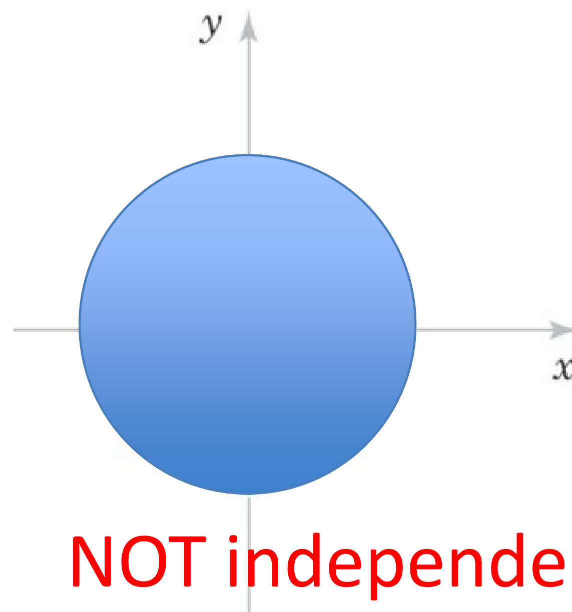
- If *X* and *Y* are independent random variables,

  $\sigma_{XY} = \rho_{XY} = 0$ (5-17)

- $\rho_{XY} = 0$ is necessary, but not a sufficient condition for independence.



Independent
covariance=0



NOT independent
covariance=0

Correlation is "normalized covariance"

- Also called:
  Pearson correlation coefficient

  $\rho_{XY} = \sigma_{XY} / \sigma_X \sigma_Y$
  is the covariance normalized to be $-1 \leq \rho_{XY} \leq 1$

Karl Pearson (1852– 1936)
English mathematician and biostatistician

Prove that ρ_{xy} is in $[-1, 1]$

$$Z_x = \frac{X - \mu_x}{\sigma_x} \quad ; \quad Z_y = \frac{Y - \mu_y}{\sigma_y}$$

$$0 \leq E((Z_x - Z_y)^2) = E(Z_x^2) + E(Z_y^2) -$$

$$- 2E(Z_x \cdot Z_y) = 2 - 2\frac{1}{\sigma_x \sigma_y} E((X - \mu_x)(Y - \mu_y)) =$$

$$2 - 2\rho_{xy} \Longrightarrow \boxed{\rho_{xy} \leq 1}$$

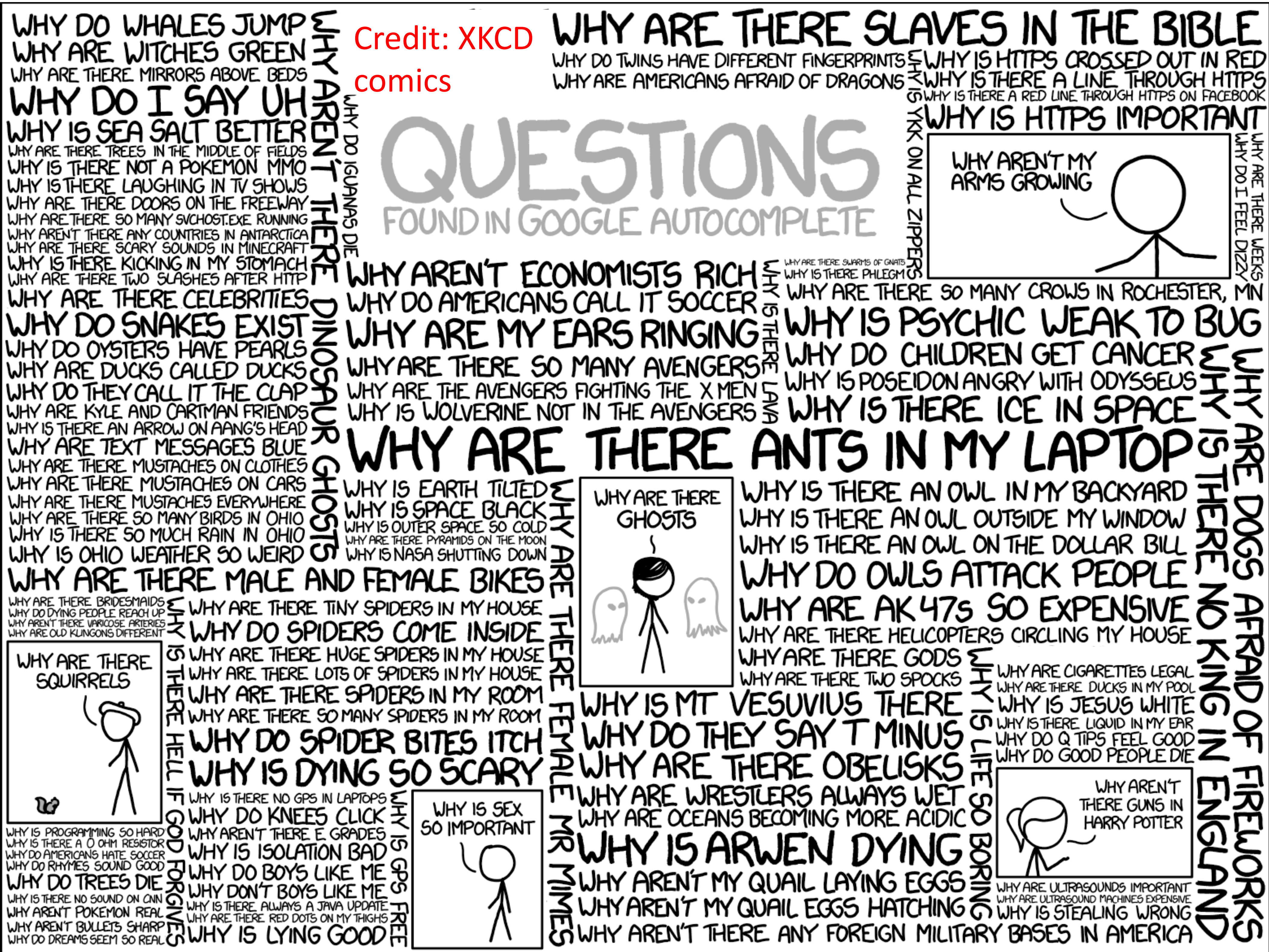
$$0 \leq E((Z_x + Z_y)^2) = E(Z_x^2) + E(Z_y^2) +$$

$$+ 2E(Z_x \cdot Z_y) = 2 + 2\rho_{xy} \longrightarrow$$

$$\Longrightarrow \boxed{\rho_{xy} \geq -1}$$

Spearman rank correlation

- Pearson correlation tests for linear relationship between X and Y
- Unlikely for variables with broad distributions → non-linear effects dominate
- Spearman correlation tests for any monotonic relationship between X and Y
- Calculate ranks (1 to n), $r_X(i)$ and $r_Y(i)$ of variables in both samples. Calculate Pearson correlation between ranks: Spearman(X,Y) = Pearson(r_X, r_Y)
- Ties: convert to fractions, e.g. tie for 6s and 7s place both get 6.5. This can lead to artefacts.
- If lots of ties: use Kendall rank correlation (Kendall tau)



Credit: XKCD comics

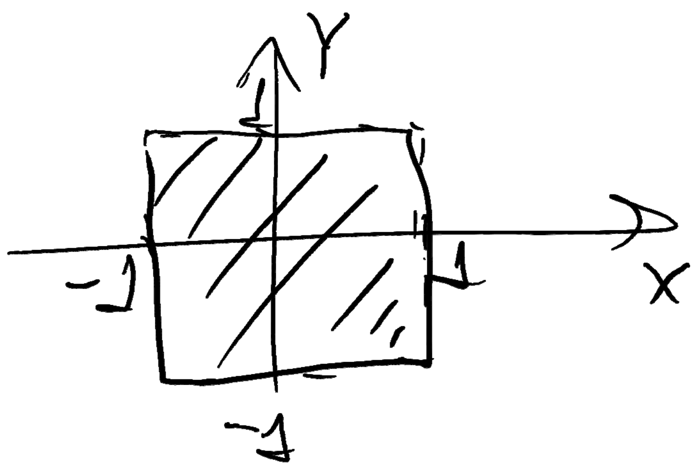
Example 3.10. Let an experiment consist of drawing a card at random from a standard deck of 52 playing cards. Define events A and B as "the card is a ♠" and "the card is a queen." Are the events A and B independent? By definition, $P(A \cdot B) = P(Q♠) = \frac{1}{52}$. This is the product of $P(♠) = \frac{13}{52}$ and $P(Q) = \frac{4}{52}$, and events A and B in question are independent. In this situation, intuition provides no help. Now, pretend that the 2♡ is drawn and excluded from the deck prior to the experiment. Events A and B become dependent since

$$\mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{13}{51} \cdot \frac{4}{51} \neq \frac{1}{51} = \mathbb{P}(A \cdot B).$$

Example 1:

Uniform distribution in the square

$-1 < X < 1, \quad -1 < Y < 1$



$$f_{XY}(x, y) = \begin{cases} c & \text{if } -1 < x < 1 \text{ and } -1 < y < 1 \\ 0 & \text{- otherwise} \end{cases}$$

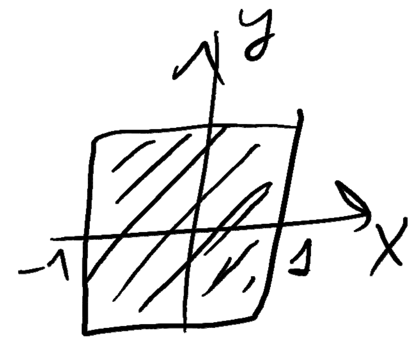
$$1 = \int_{\text{square}} dx\,dy\, f_{XY}(x,y) = c \cdot \text{Area} = c \cdot 4 \Rightarrow c = \frac{1}{4}$$

Are X and Y independent? Yes they are

Let's test if $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy =$$

$$= \int_{-1}^{1} \frac{1}{4}\,dy = \frac{1}{2} \text{ if } -1 < x < 1$$



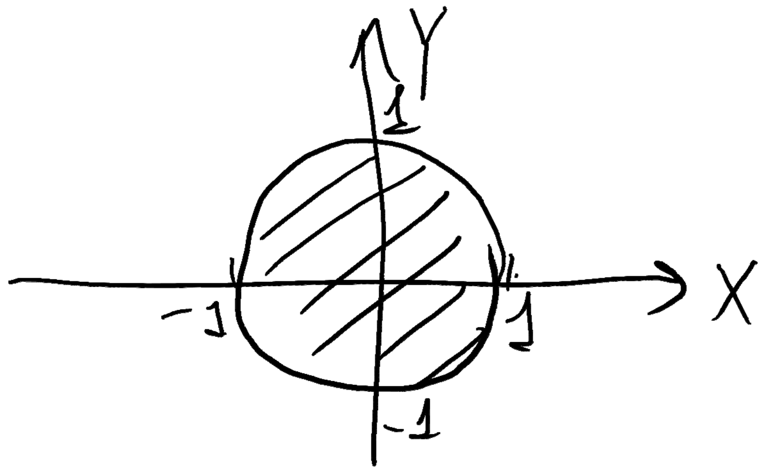
Same for $f_Y(y) = \frac{1}{2}$ if $-1 < y < 1$

$$\frac{1}{4} = f_{XY}(x,y) = \frac{1}{2} \cdot \frac{1}{2} = f_X(x) \cdot f_Y(y)$$

if both x & y are in $[-1,1]$

0 otherwise

X and Y are uniformly distributed in the disc $x^2+y^2\leq 1$



Are they independent?

A. yes
B. no
C. I could not figure it out

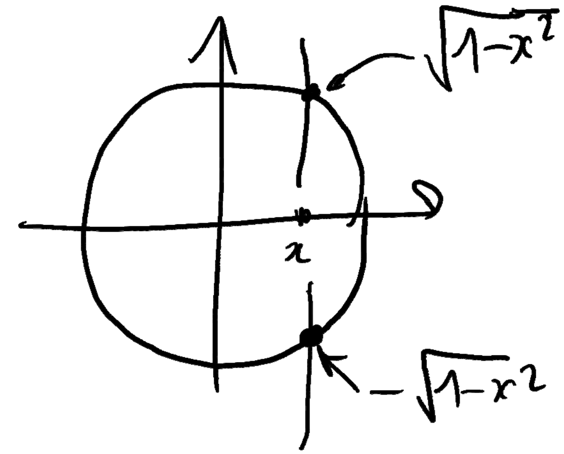
Get your i-clickers

Joint PDF $f_{XY}(x,y) = \frac{1}{\text{area}} = \frac{1}{\pi}$ if x, y in the disc

0 – otherwise

Marginal distributions:

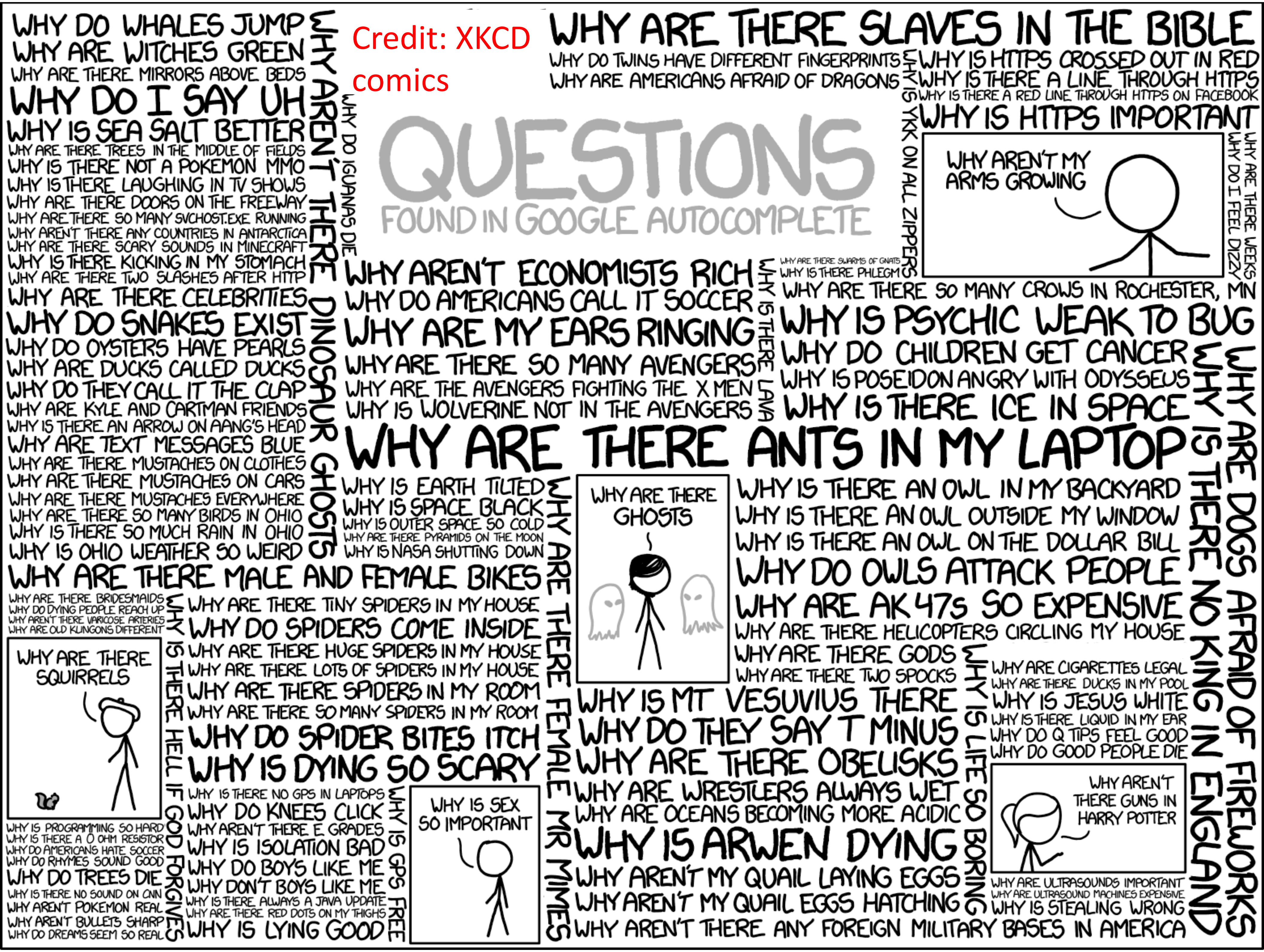
$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{dy}{\pi} = \frac{2\sqrt{1-x^2}}{\pi}$$



Same for $f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi}$

$$\frac{1}{\pi} = f_{XY}(x,y) \neq \frac{2}{\pi}\sqrt{1-x^2} \cdot \frac{2}{\pi}\sqrt{1-y^2} = f_X(x) \cdot f_Y(y)$$

Variables are **NOT** independent



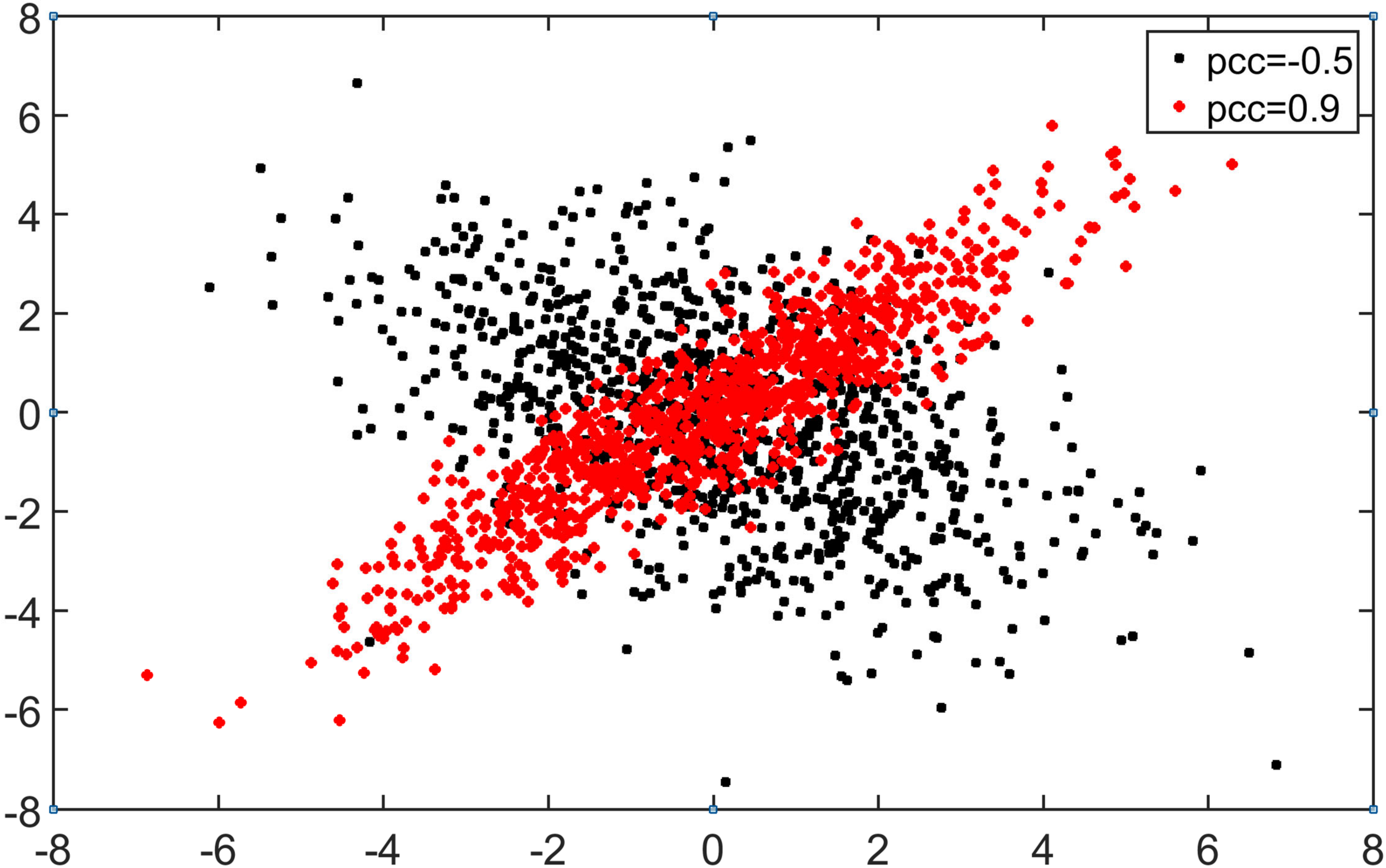
Credit: XKCD comics

Matlab exercise: Correlation/Covariation

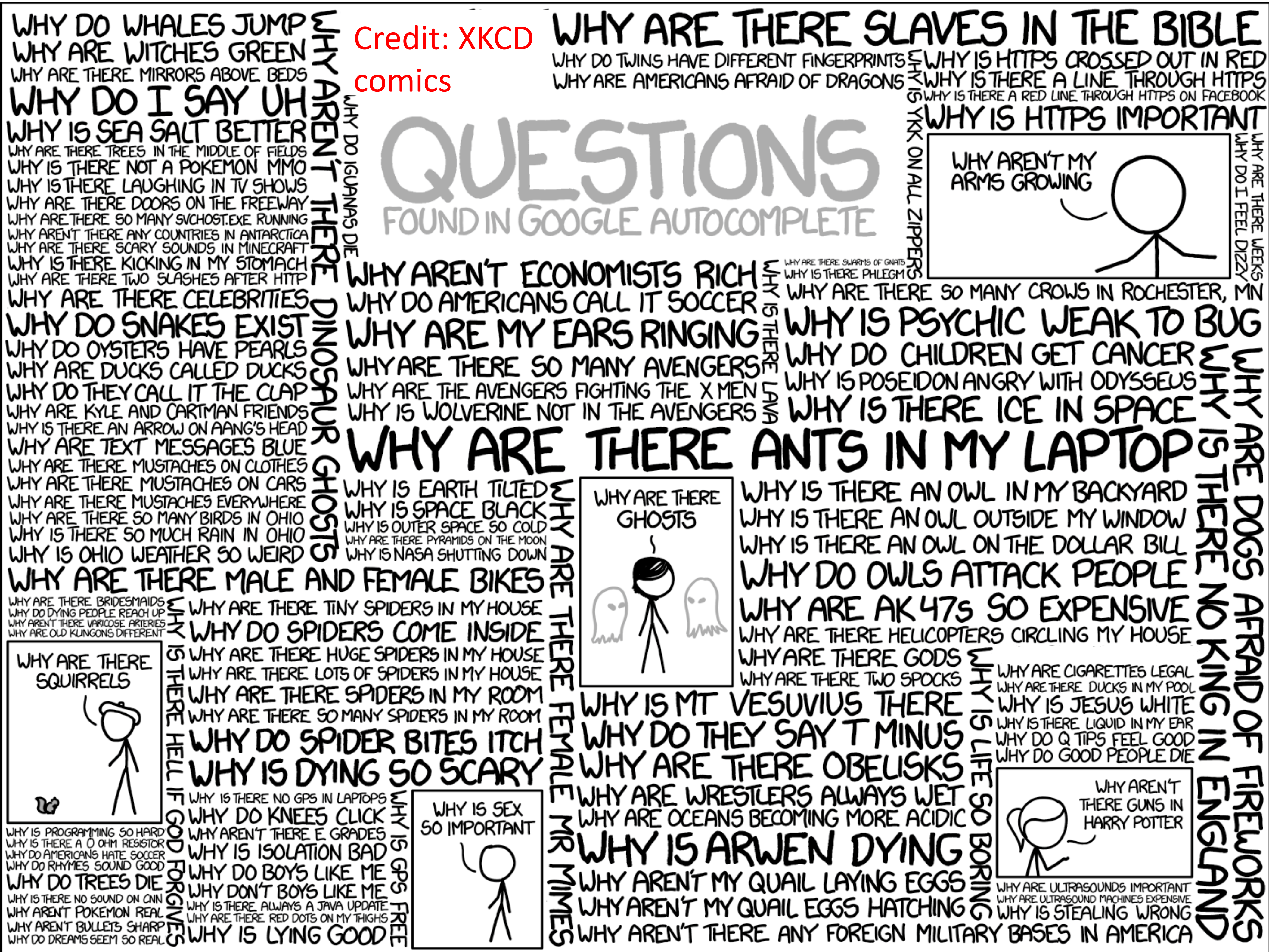
- Generate a sample with Stats=100,000 of two Gaussian random variables r1 and r2 which have mean 0 and standard deviation 2 and are:
  - Uncorrelated
  - Correlated with correlation coefficient 0.9
  - Correlated with correlation coefficient -0.5
  - Trick: first make uncorrelated r1 and r2. Then make anew variable: r1mix=mix.*r2+(1-mix.^2)^0.5.*r1; where mix= corr. coeff.
- For each value of mix calculate covariance and correlation coefficient between r1mix and r2
- In each case make a scatter plot: plot(r1mix,r2,'k.');

Matlab exercise: Correlation/Covariation

```
1.  Stats=100000;
2.  r1=2.*randn(Stats,1);
3.  r2=2.*randn(Stats,1);
4.  disp('Covariance matrix='); disp(cov(r1,r2));
5.  disp('Correlation=');disp(corr(r1,r2));
6.  figure; plot(r1,r2,'k.');
7.  mix=0.9; %Mixes r2 to r1 but keeps same variance
8.  r1mix=mix.*r2+sqrt(1-mix.^2).*r1;
9.  disp('Covariance matrix='); disp(cov(r1mix,r2));
10. disp('Correlation=');disp(corr(r1mix,r2));
11. figure; plot(r1mix,r2,'k.');
12. mix=-0.5; %REDO LINES 8-11
```



pcc=-0.5
pcc=0.9
8
6
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4
6
8



Credit: XKCD comics