Stat310

Bivariate rvs continued

Hadley Wickham

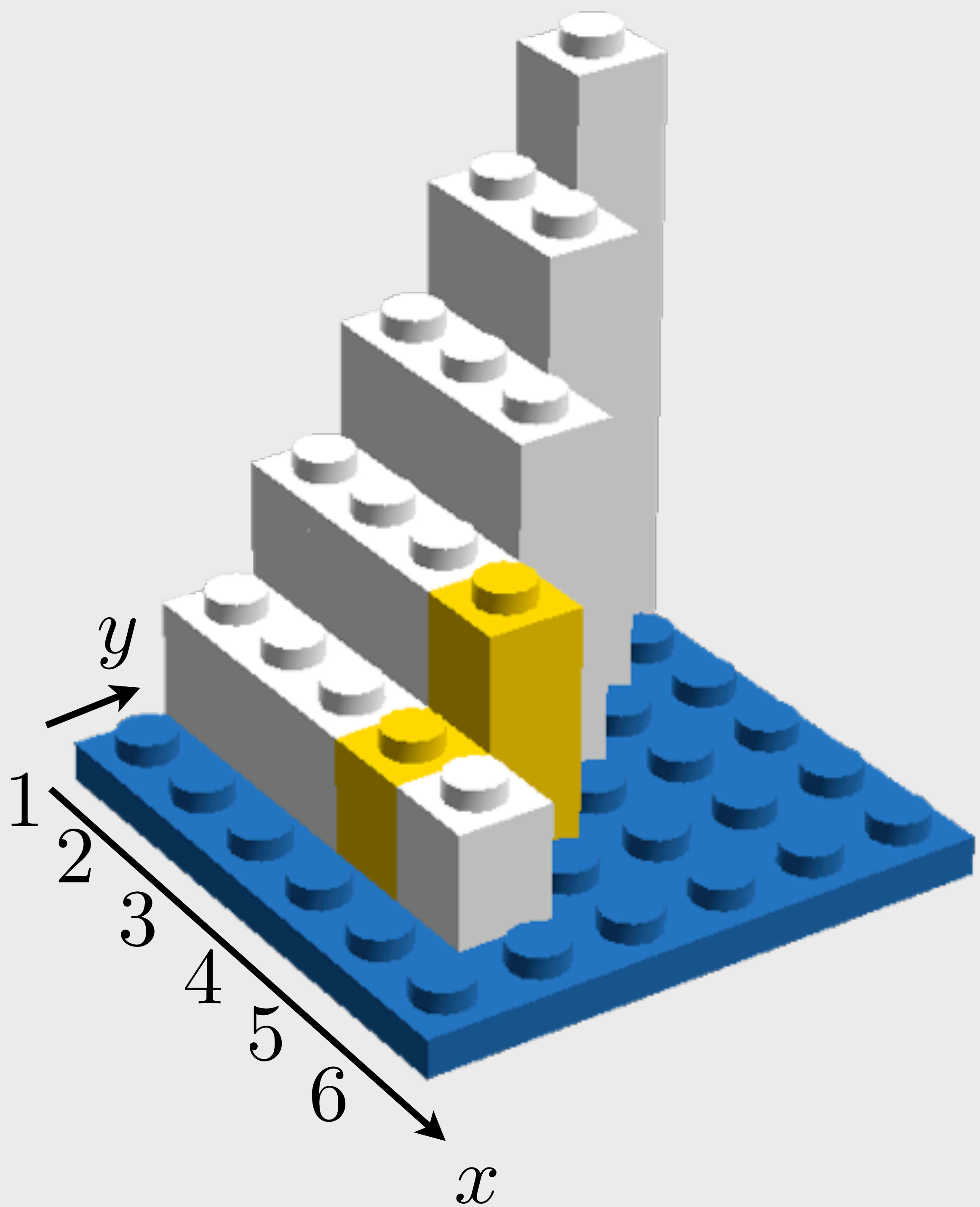
Assessment

- Tests due 5pm in homework box
- HW 5 back
- No homework over spring break (just your stats in practice)

1. Marginal and conditional distributions
2. Expectation
3. Covariance and correlation

Warm up

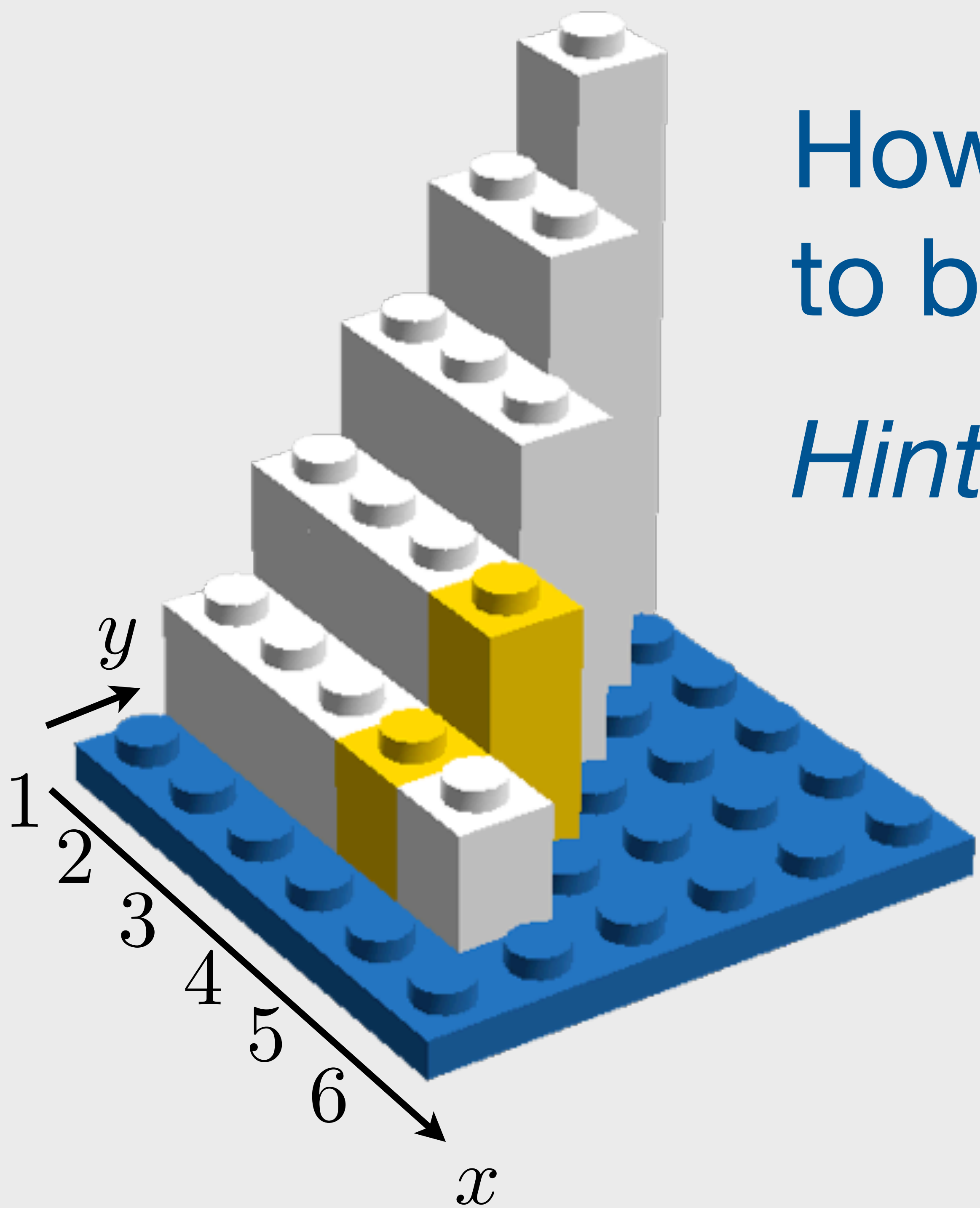
Warm Up 1



What is P(X = 4)?

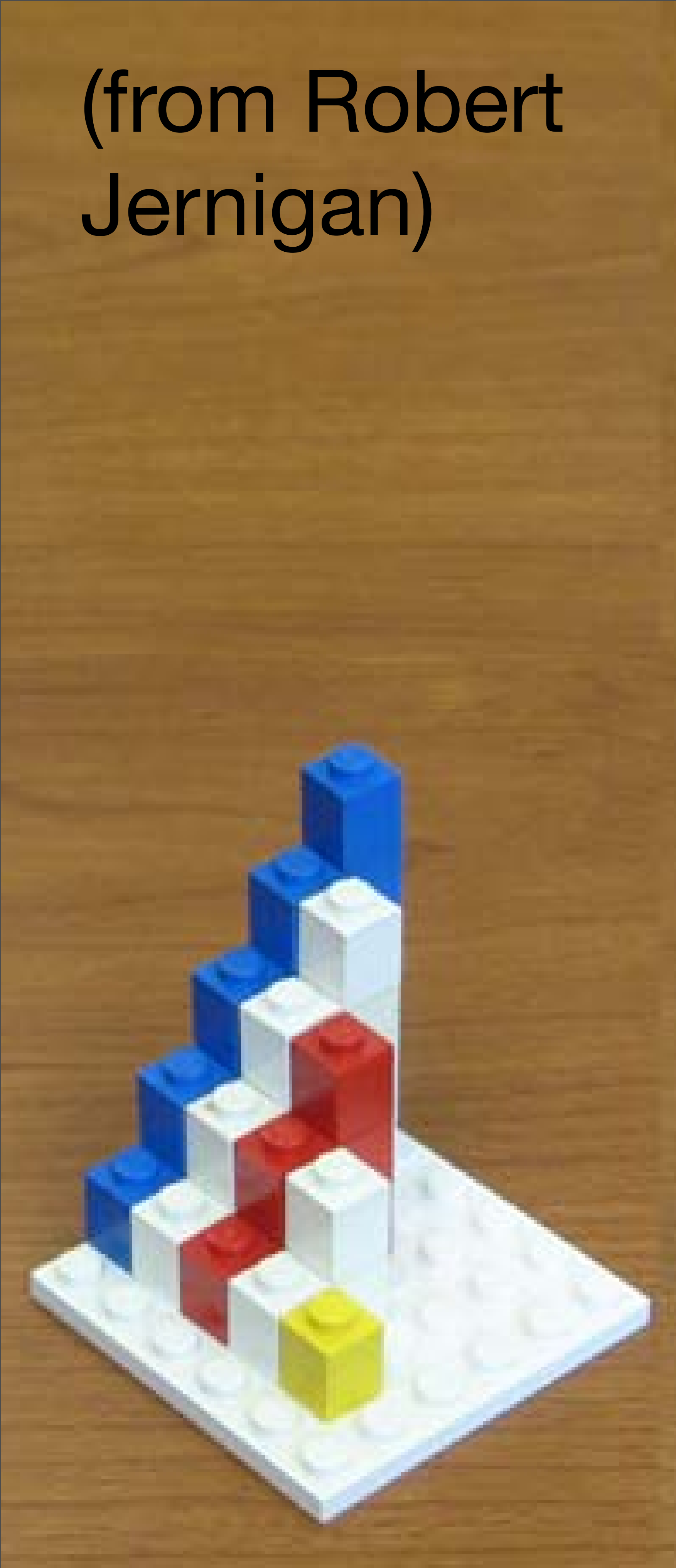
Hint: There are 35 legos total

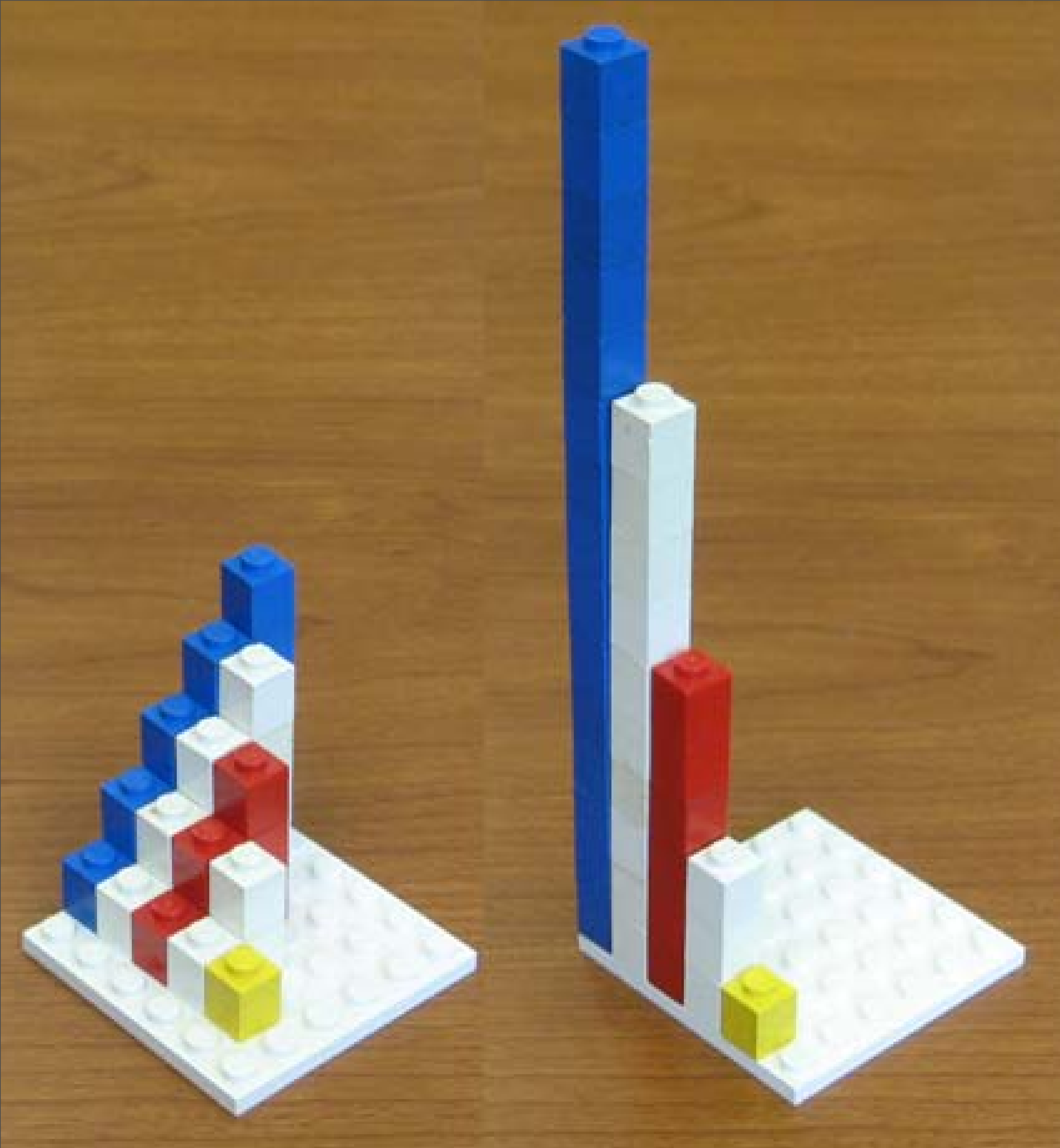
Warm Up 2

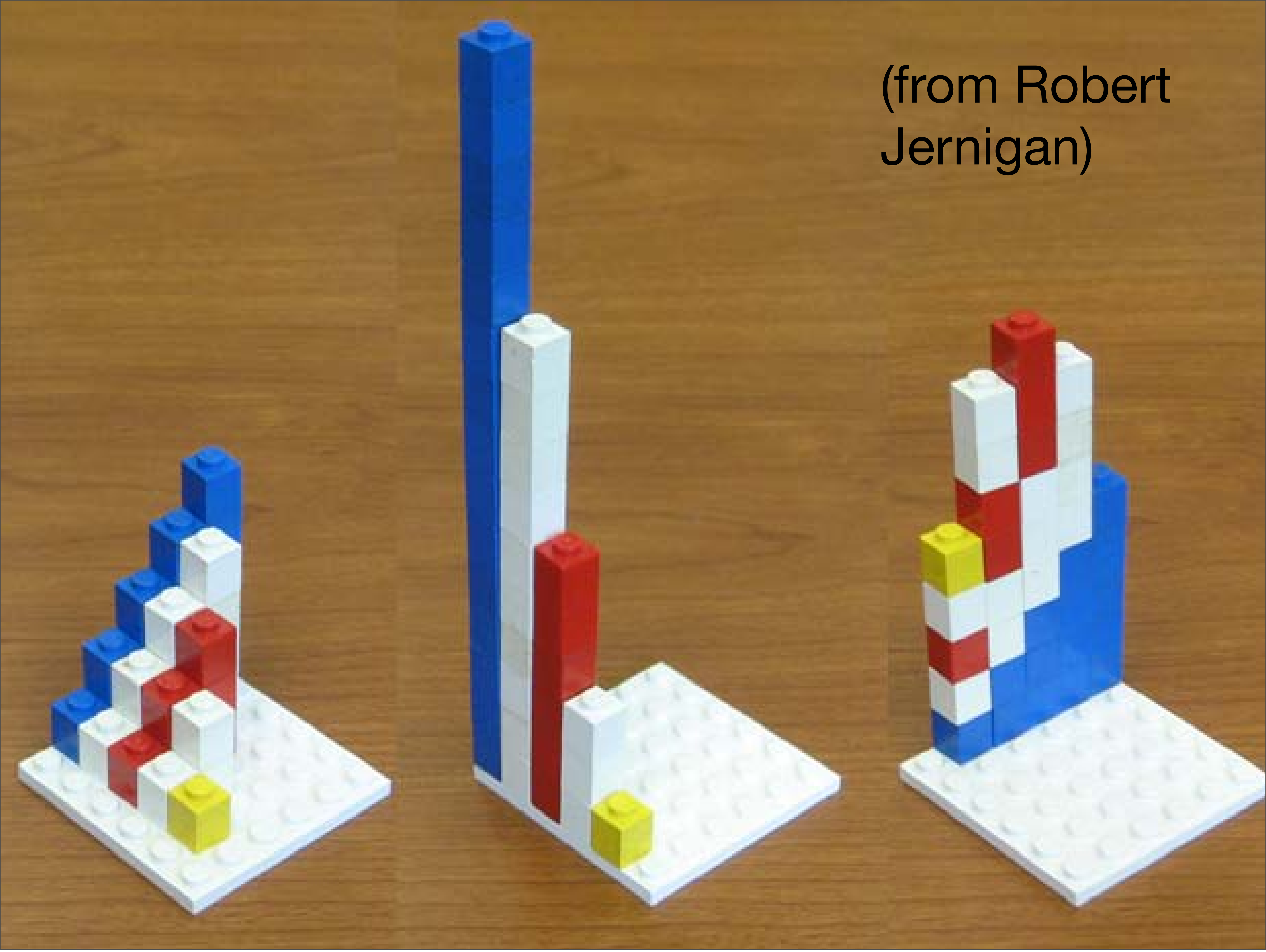


How can we use this *pdf* to build a 1D *pdf* for X?

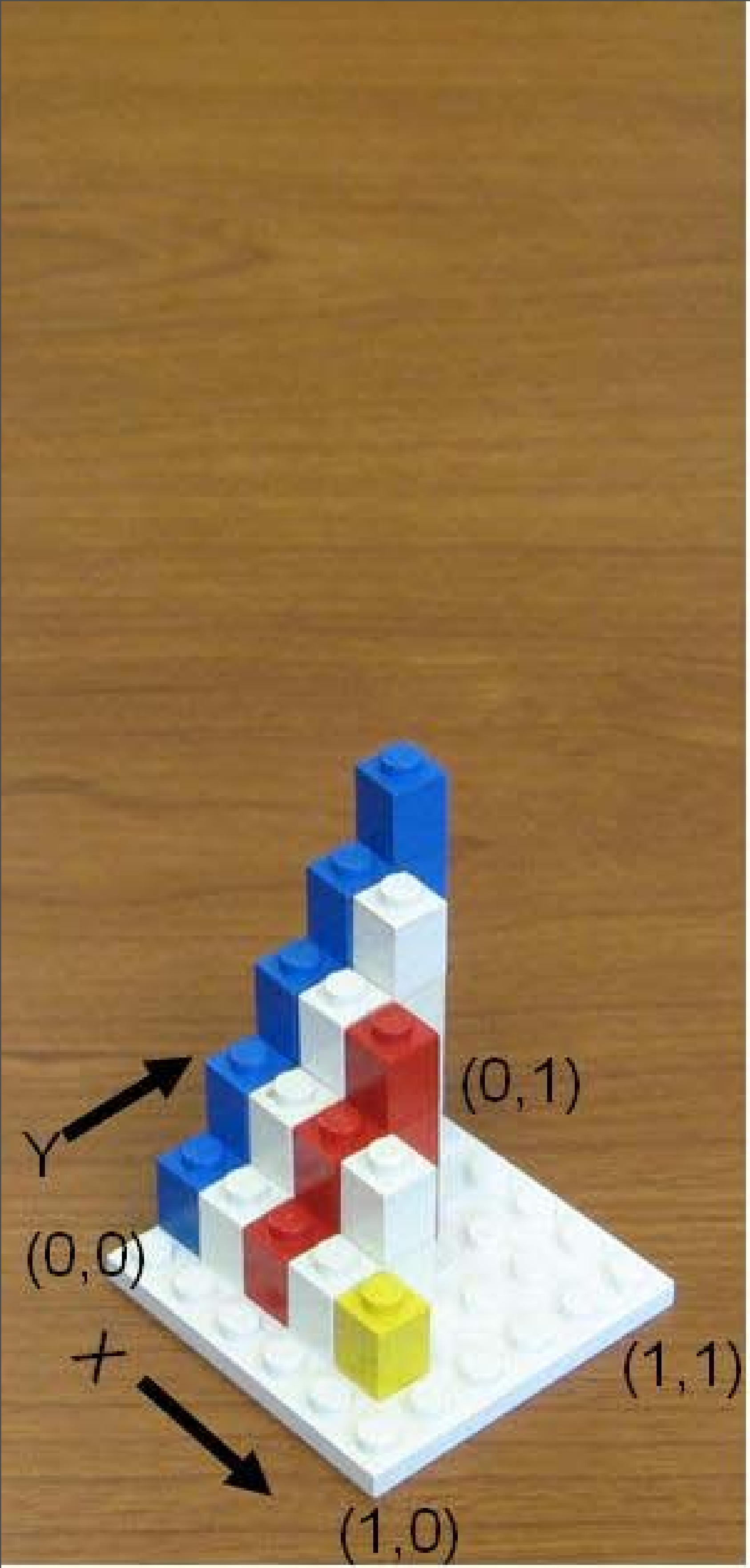
Hint: f(X = 4) * 35 = 3


(from Robert
Jernigan)

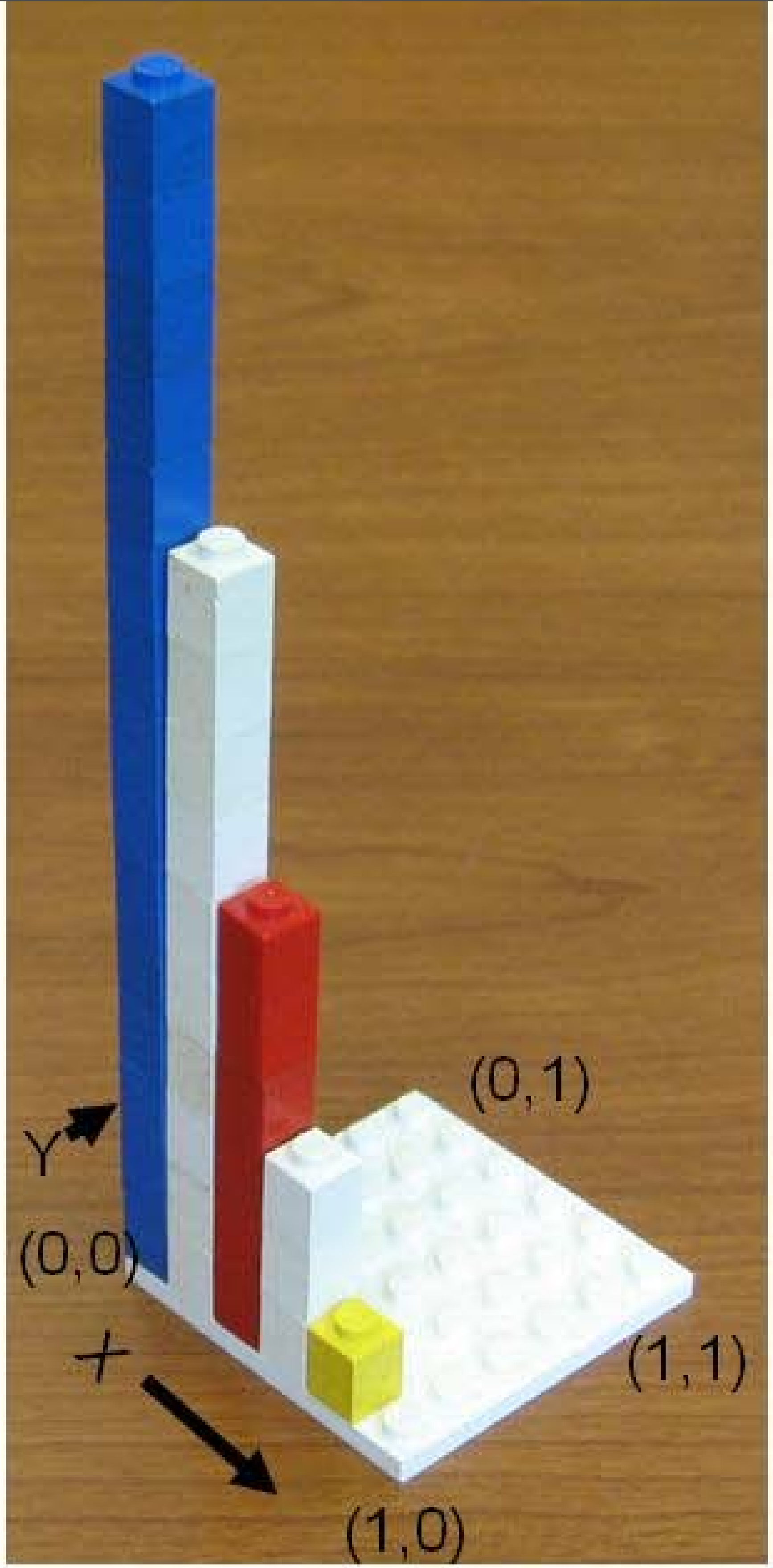



(from Robert
Jernigan)

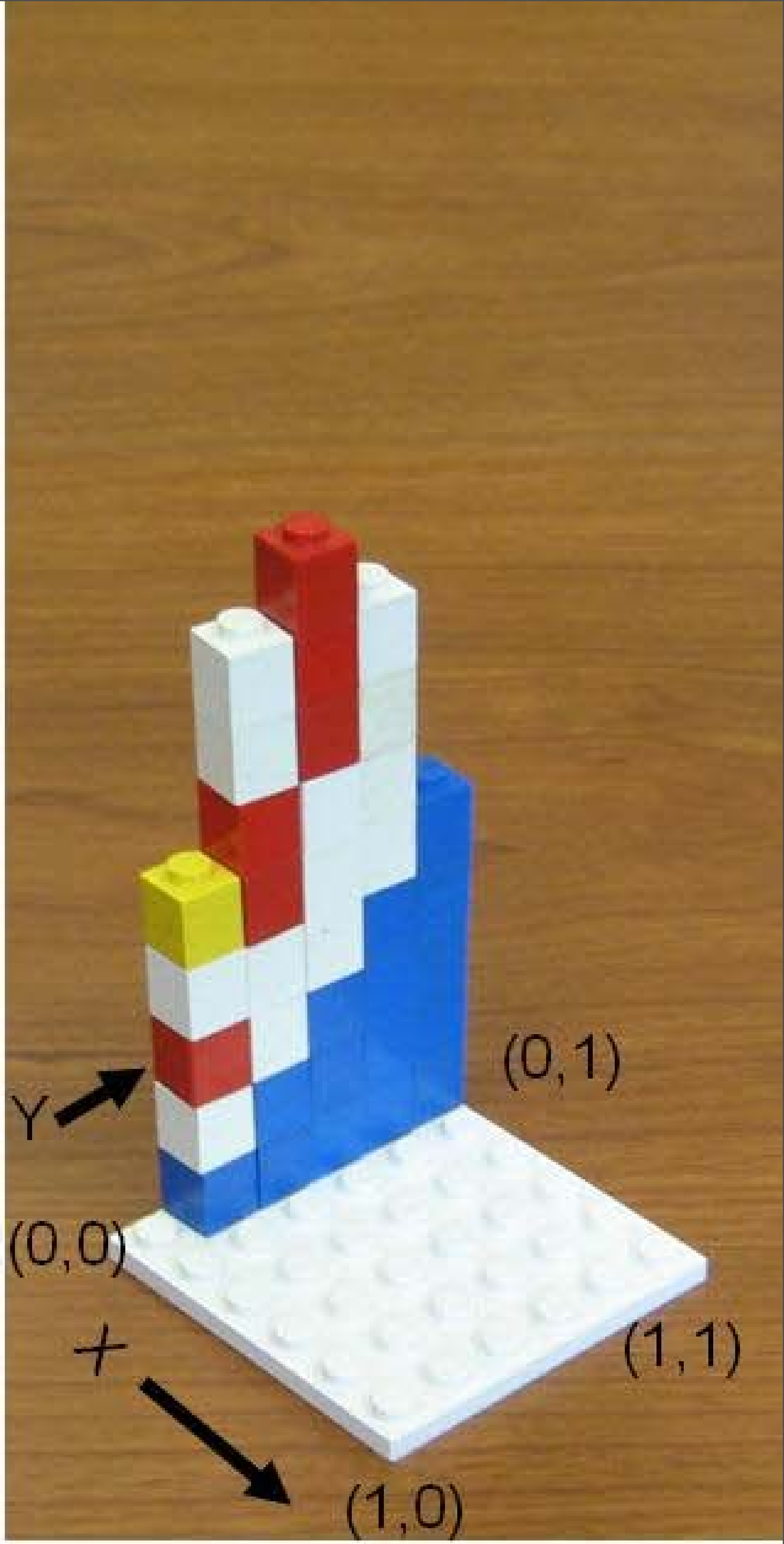
Marginal & Conditional



$$f(x,y)=\begin{cases}6y & 0<x<1-y,\ 0<y<1\\ 0 & otherwise\end{cases}$$



$$f_1(x)=3(1-x)^2,\ \text{for } 0<x<1;$$
$$=0 \text{ otherwise}$$



$$f_2(y)=6y(1-y),\ \text{for } 0<y<1$$
$$=\ 0 \text{ otherwise}$$

Marginal Distributions

Given a joint distribution $f(x, y)$, we can calculate the individual pdfs for x and y. These are known as the "marginal distributions." They are denoted by

$$f_X(x) \quad \& \quad f_Y(y)$$

Marginal distributions (continuous case)

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy$$

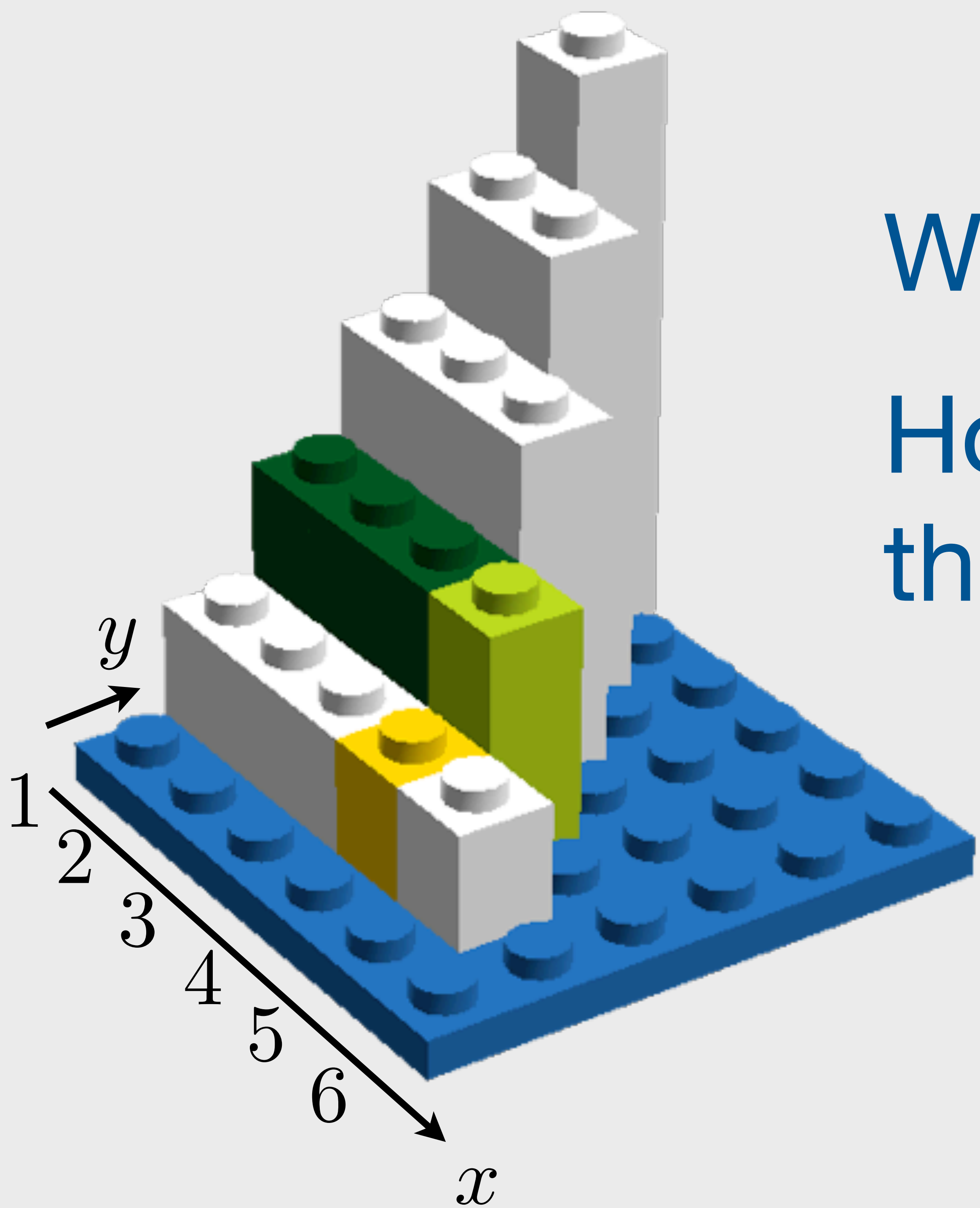
$$f_Y(y) = \int_{\mathbb{R}} f(x, y) dx$$

Marginal distributions (discrete case)

$$f_X(x) = \sum_{y \in S} f(x, y)$$

$$f_Y(y) = \sum_{x \in S} f(x, y)$$

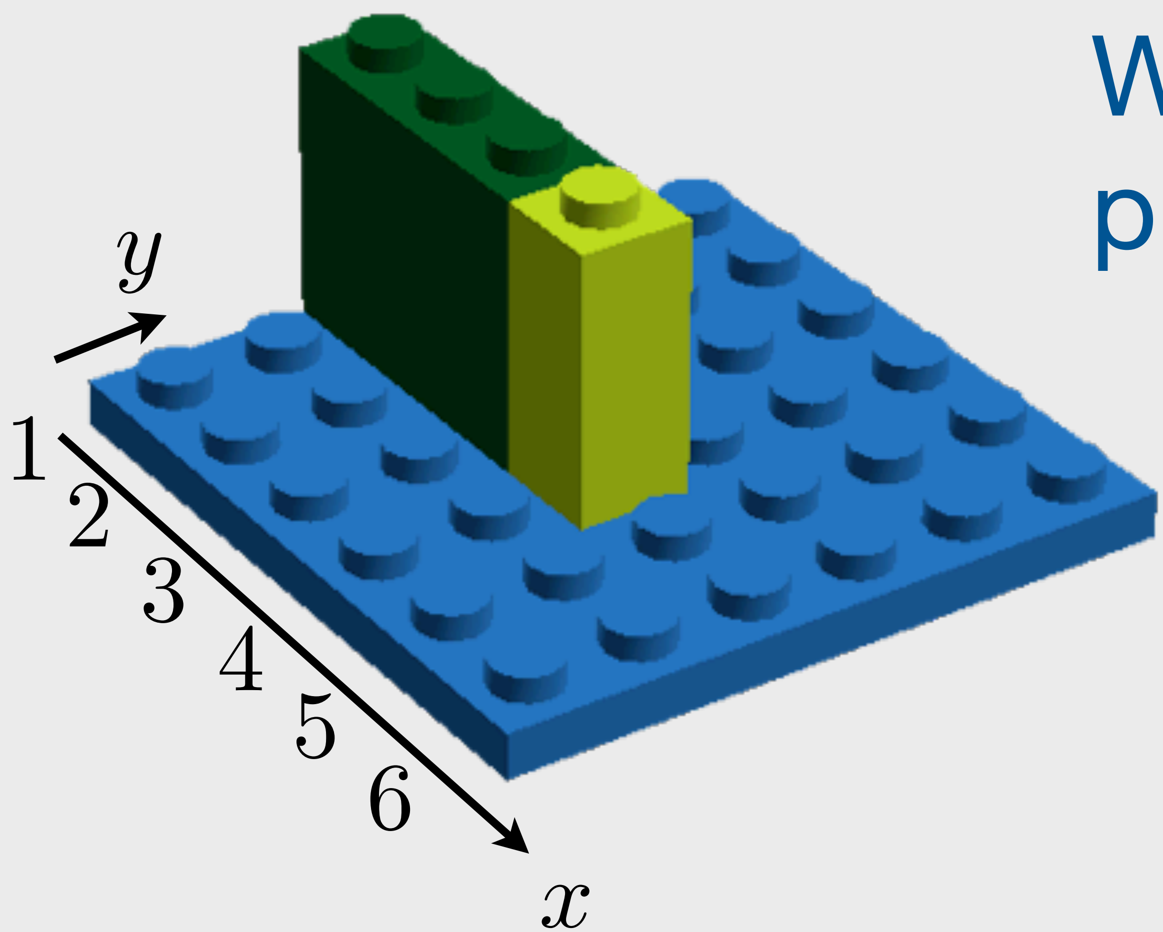
Challenge



What is P(X = 4| Y = 3)?

How do you calculate this?

Challenge



Conditioning reduces the probability of everything outside of the condition to 0.

What remains is a new probability distribution.

Conditional Distributions

By fixing the value of x or y in $f(x,y)$, we create a conditional distribution. It describes the updated probabilities for the remaining variable.

$$f_{X|Y=y}(x) \;\&\; f_{Y|X=x}(y)$$

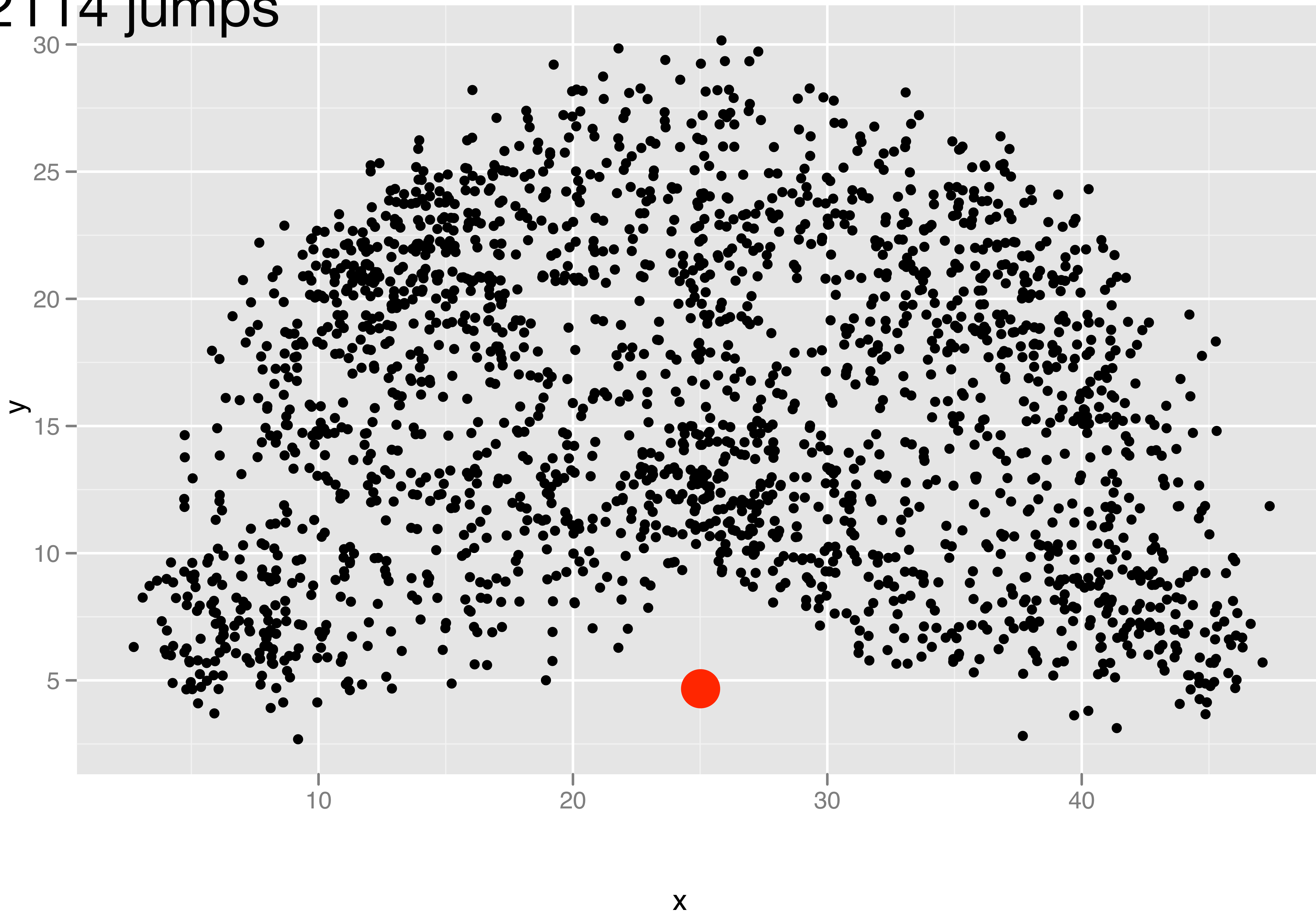
Conditional distributions

$$P(Y = y|X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)}$$

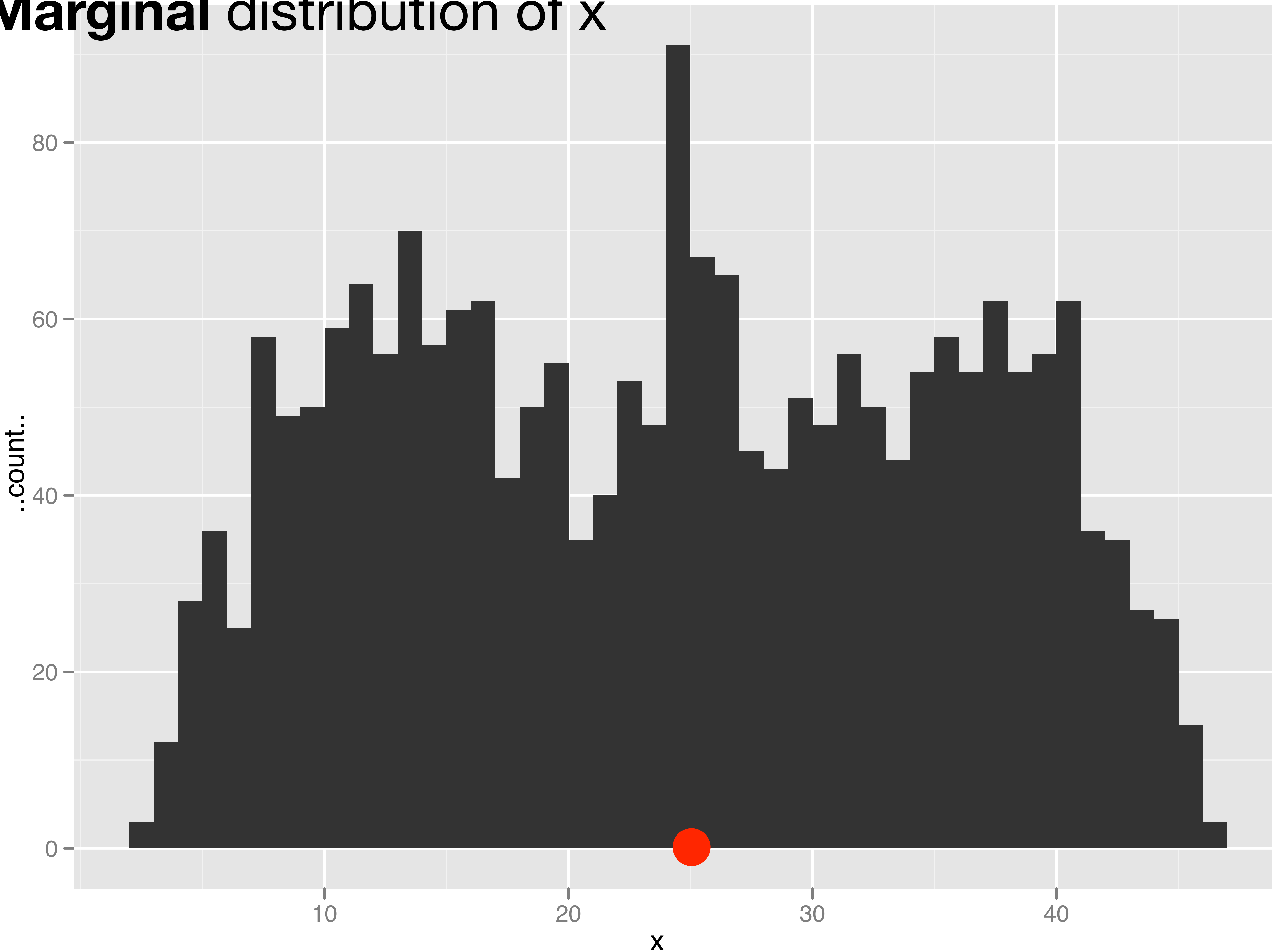
$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

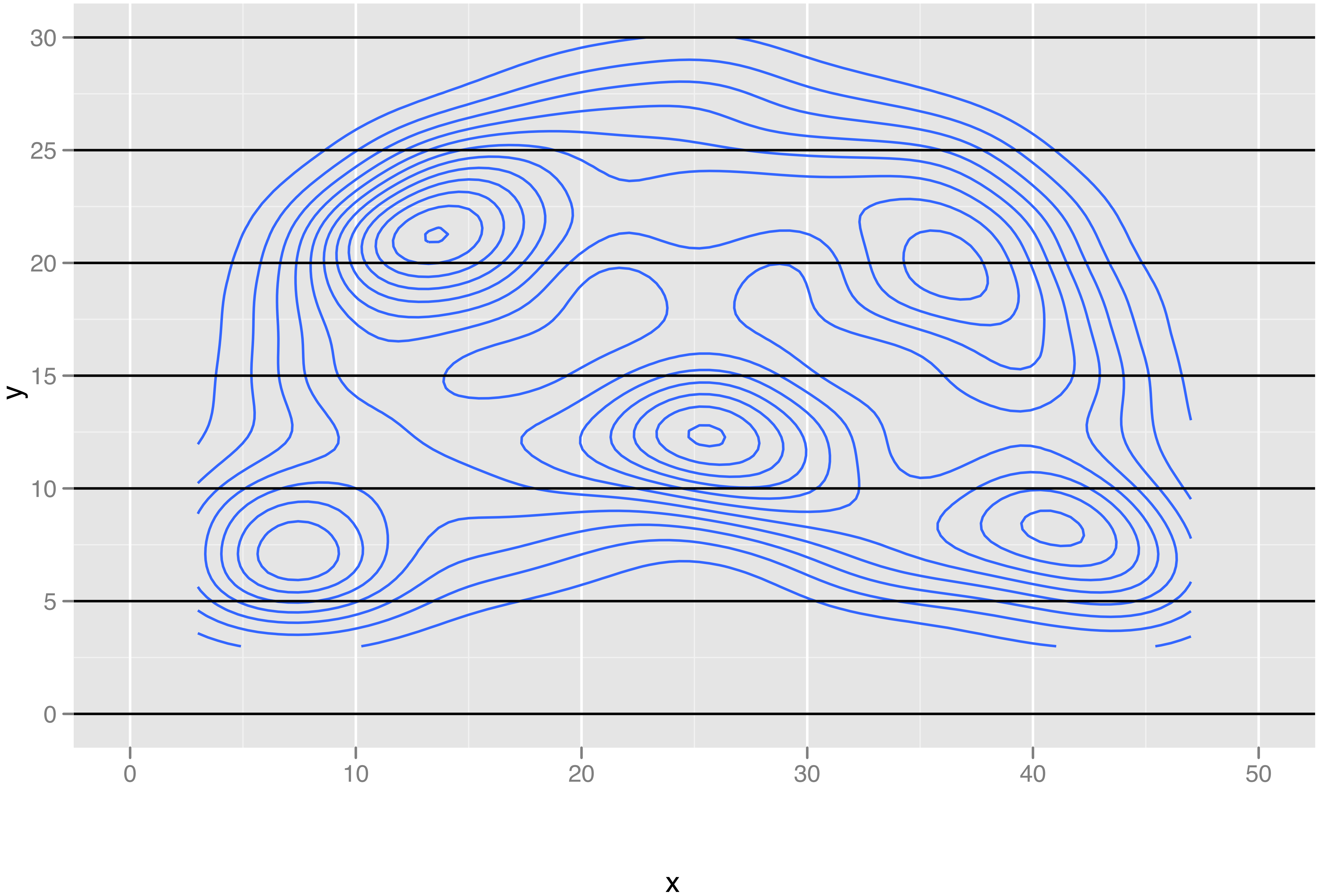
$$f_{Y|X=x}(y) \times f_X(x) = f_{X,Y}(x, y)$$

Joint distribution of x and y
2114 jumps

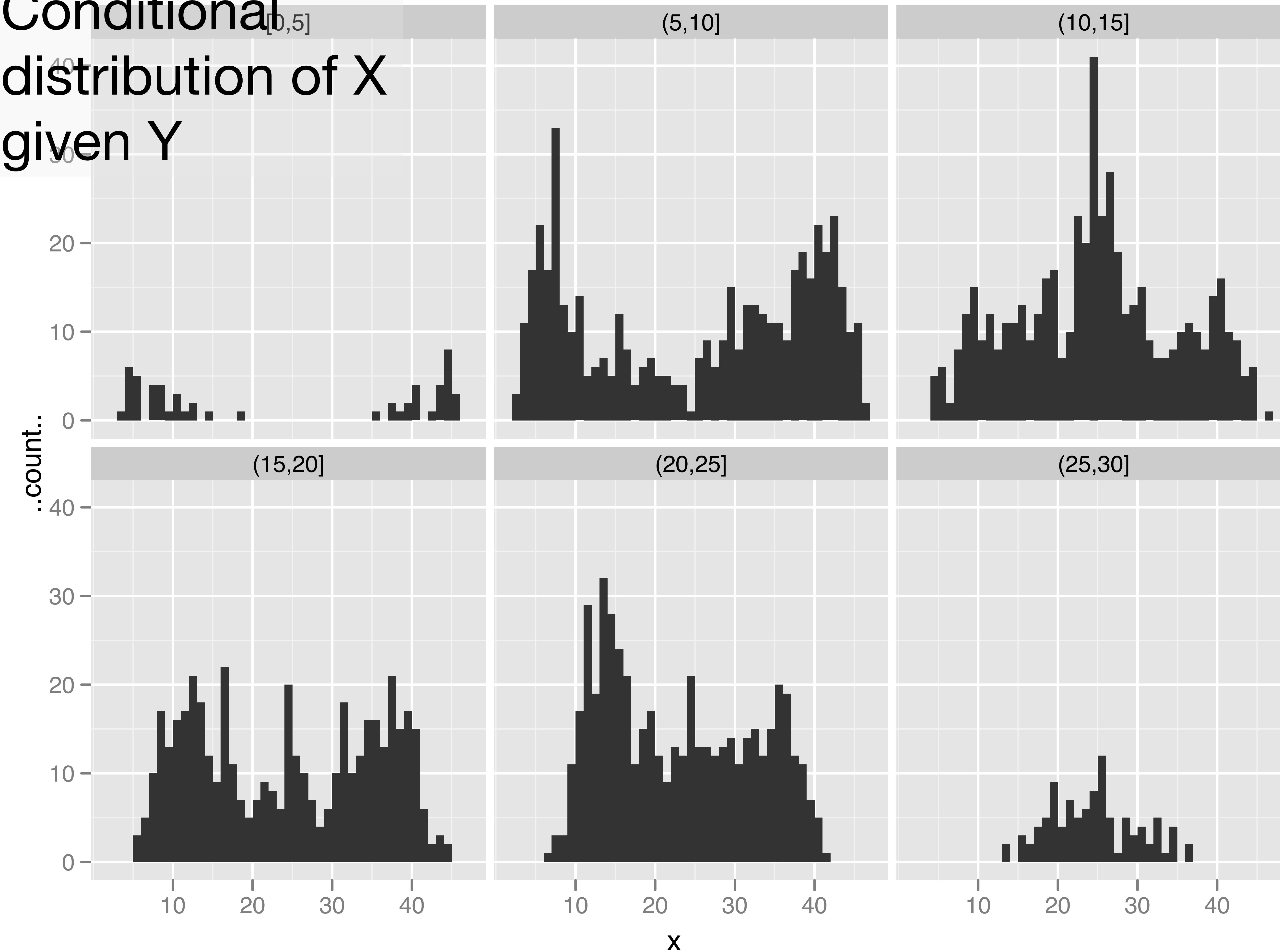


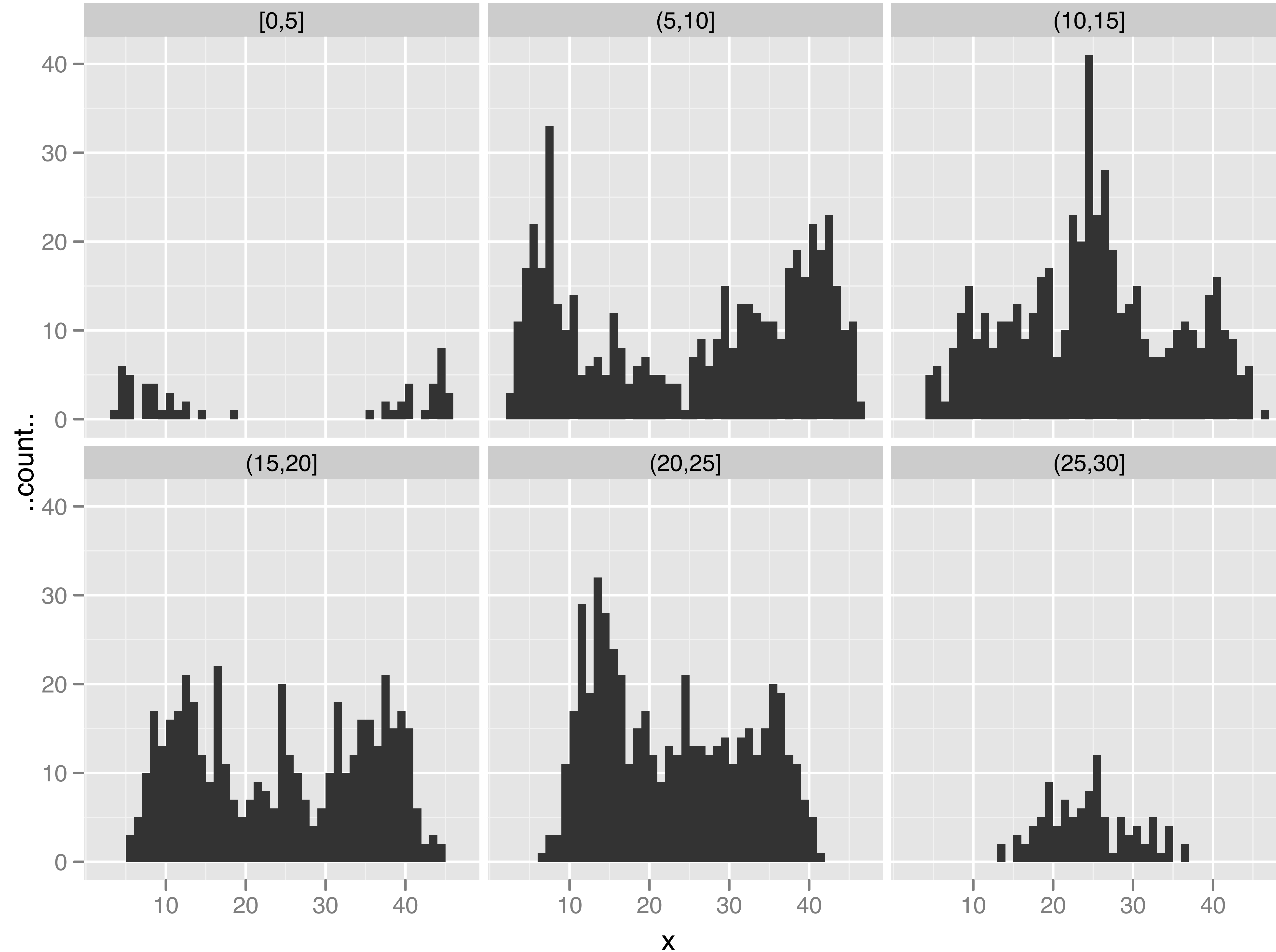
Marginal distribution of x




30
25
20
15
10
5
0
y
0
10
20
30
40
50
x

Conditional distribution of X given Y




[0,5]
(5,10]
(10,15]
(15,20]
(20,25]
(25,30]
..count..
0
10
20
30
40
10
20
30
40
x

2d 1d

Marginal

integrate

Joint

fix value,
re-normalise

Conditional

Use

We are only going to learn one **named** bivariate distribution (the bivariate normal)

Normally we will produce a joint distribution by combining a marginal distribution with a conditional distribution.

Your turn

$X \sim \text{Exponential}(\theta = 1)$

$Y \mid X = x \sim \text{Exponential}(\theta = x)$

What is the joint distribution of X and Y?

Expectation

$$E(u(X,Y)) = \iint_S u(x,y) f(x,y)\, dx\, dy$$

$$E(X) = \iint_S x f(x,y)\, dx\, dy$$

$$E(Y) = \iint_S y f(x,y)\, dx\, dy$$

Your turn

If X and Y are independent what is E(XY) ?

$$E(XY) = E(X)E(Y)$$

if X and Y are independent

$$E(aX + bY) = aE(X) + bE(Y)$$

expectation is still a linear operator!

Marginal expectation

$$E(X) = \iint_S x f(x, y)\, dx\, dy$$

$$E(Y) = \iint_S y f(x, y)\, dx\, dy$$

Conditional expectation

$$E(X|Y = y) = \int_{\mathbb{R}} f_{X|Y=y}(x)\, dx$$

$$E(Y|X = x) = \int_{\mathbb{R}} f_{Y|X=x}(y)\, dy$$

Your turn

A bird lays on average 3 eggs per season. For each egg, P(egg hatches) = 0.7.

What is the probability that more than 3 eggs will hatch in a season?

What is the expected number of hatched egg per season?

Independenc
e

Independence

$P(X \mid Y) = X$

$P(Y \mid X) = Y$

$\Rightarrow f(x, y) = f(x)\, f(y),\ S = S_x \times S_y$

Recall,

$$f_{Y|X=x}(y) \times f_X(x) = f_{X,Y}(x,y)$$

If x and y are independent, what does

$$f_{Y|X=x}(y) = ?$$

Hint: Independence implies that

P(X|Y) = X

P(Y|X) = Y

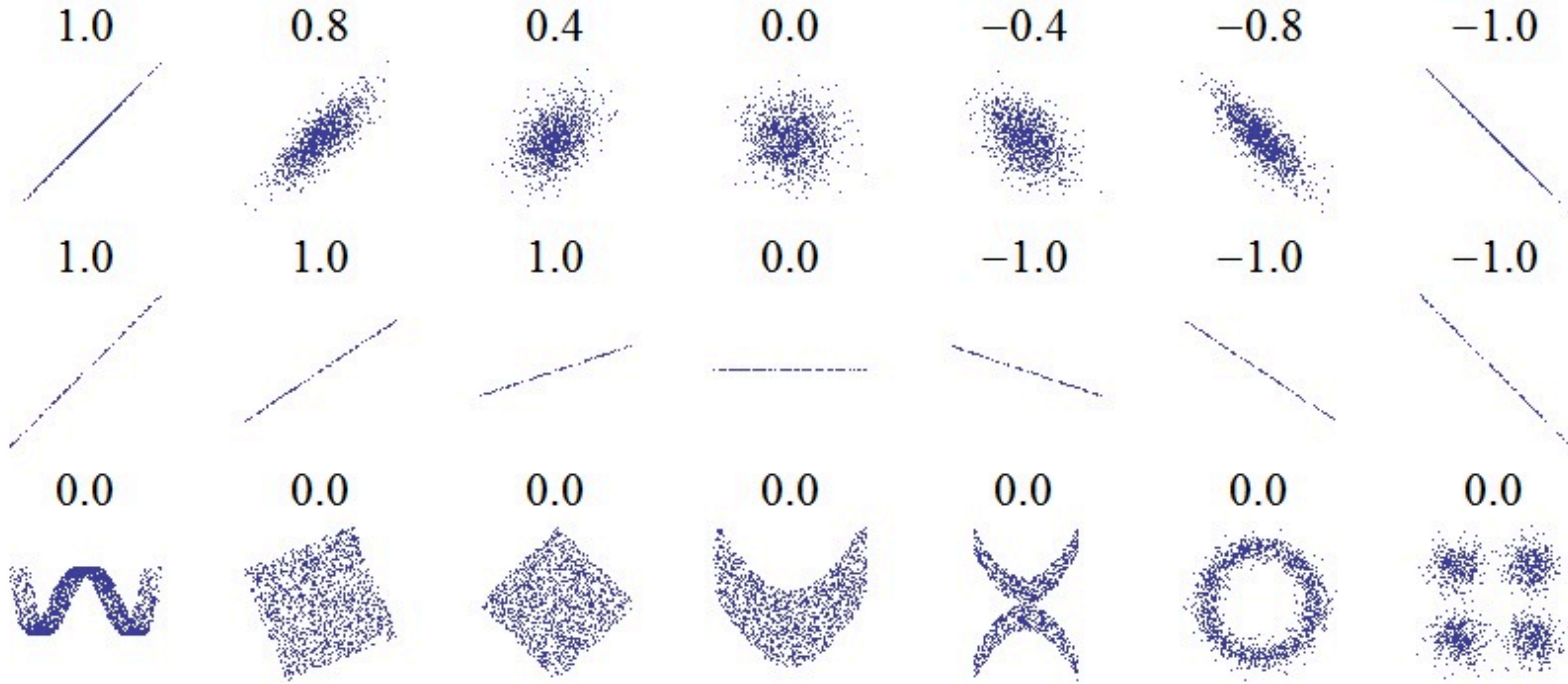
Dependence

Only one way for rv's to be independent.

Many ways to be dependent. Useful to have some measurements to summarise common forms of dependence.

Later we'll talk about one measure of dependence: correlation.

Correlation



From: http://en.wikipedia.org/wiki/File:Correlation_examples.png

ρ

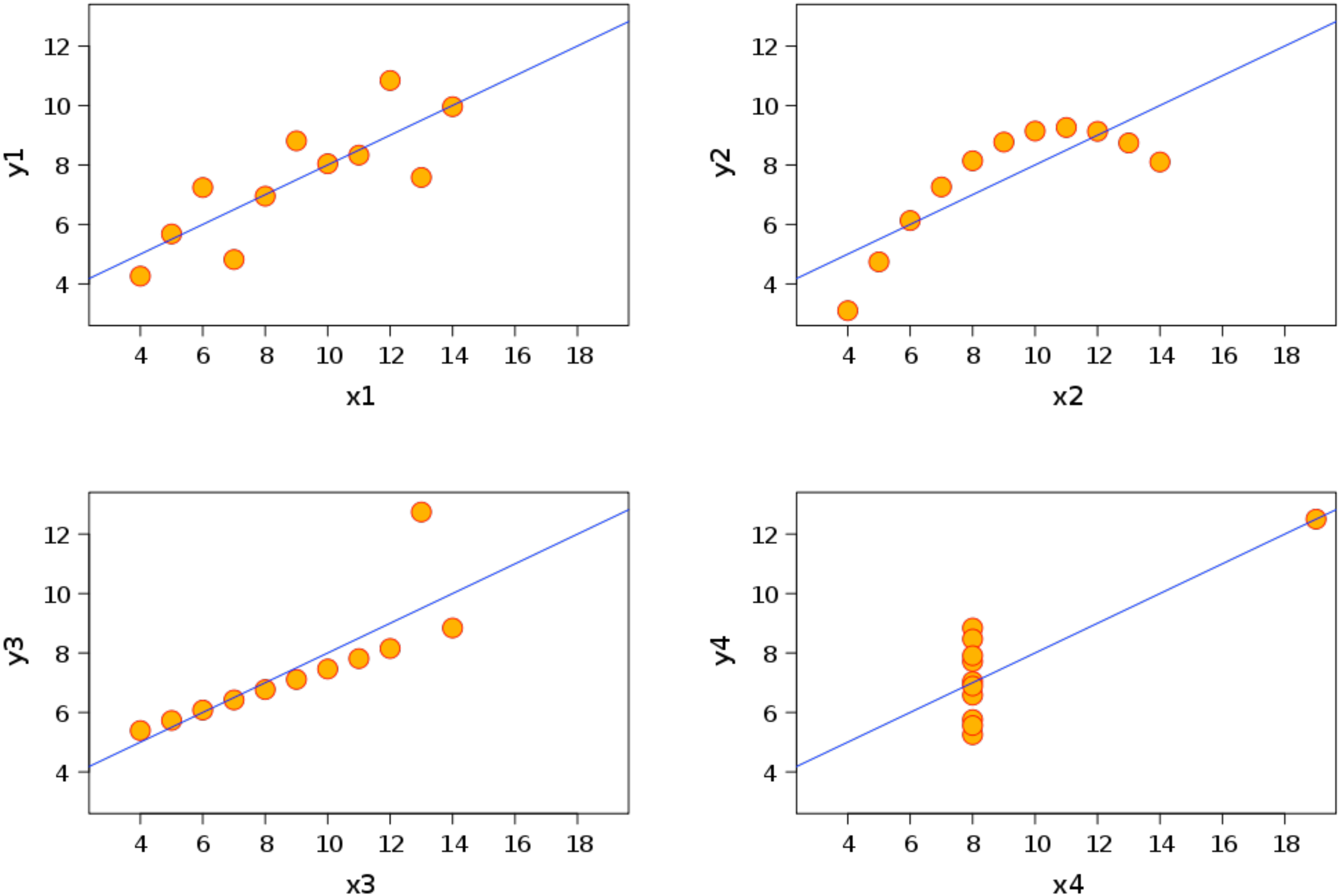
What is the range of ρ?

For what value of ρ are X and Y most strongly positively related?

For what value of ρ are X and Y most strongly negatively related?

For what value of ρ is there no (linear) relationship between X and Y?

Correlation = 0.8



Covariance

Easiest to define correlation in terms of another function: co**variance**

$$\mathrm{Cov}(X, Y) = E[\ (X - E(X))(Y - E(Y))\ ] = \sigma_{XY}$$

$$\mathrm{Cov}(X, X) = \mathrm{Var}(X) = \sigma_{XX} = \sigma_X^2$$

Covariance

If $\sigma_{XY} > 0$, then X tends to increase when Y increases

If $\sigma_{XY} < 0$, then X tends to decrease when Y increases

If $\sigma_{XY} = 0$, then there is no linear relationship between X and Y
(but there may be a non-linear relationship!)

Alternative

Is there another way to compute the covariance? (Think about the two ways of computing the variance)

If X and Y are independent,
what is Cov(X, Y)?

Counterexample

Let X be a random variable with $E(X) = 0$, $E(X^2) = 10$, $E(X^3) = 0$. Let $Y = X^2$.

Are X and Y independent?

Are X and Y uncorrelated?

Correlation

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_{XX}\sigma_{YY}}}$$