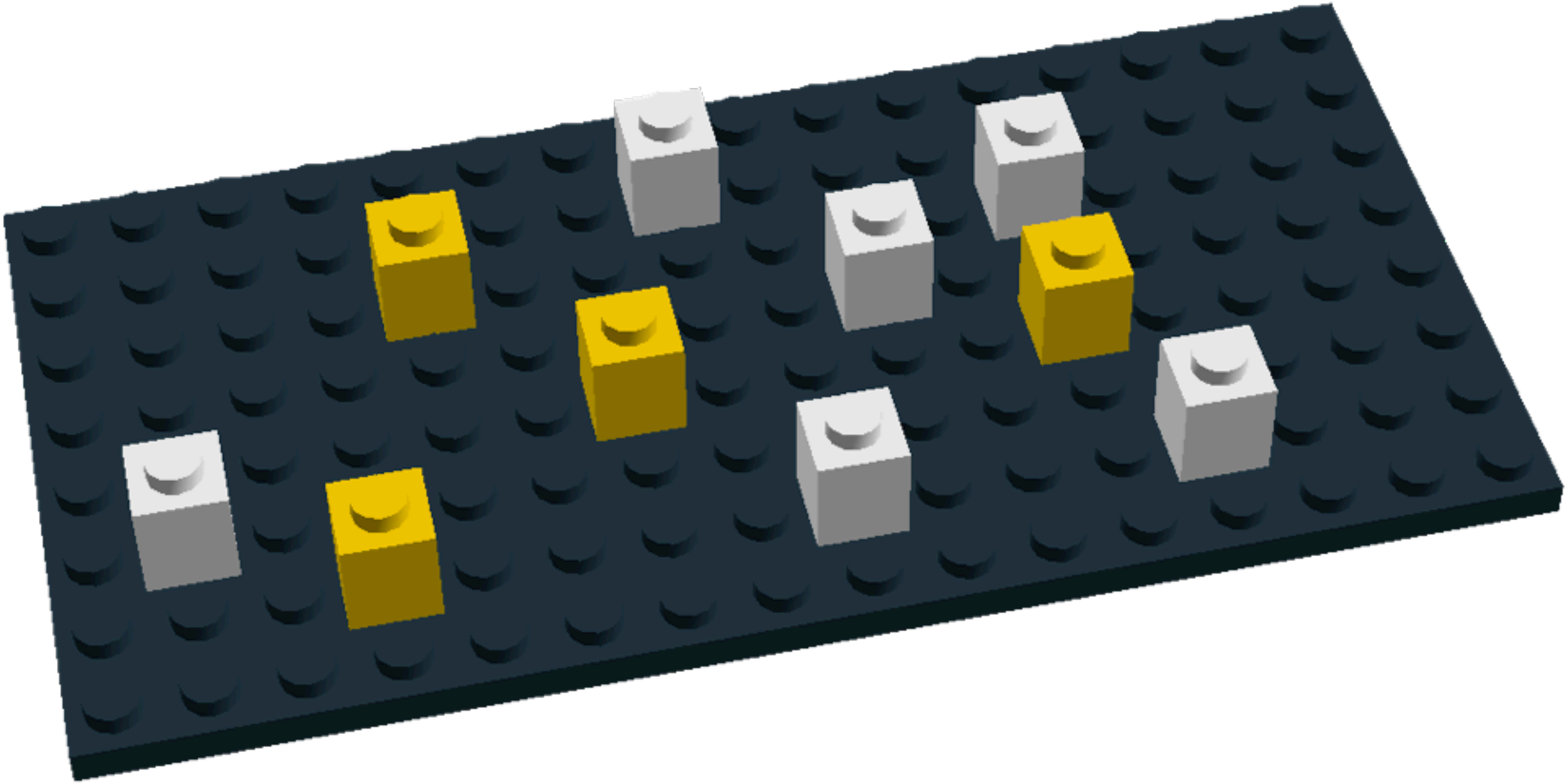
Stat310

Bivariate random variables

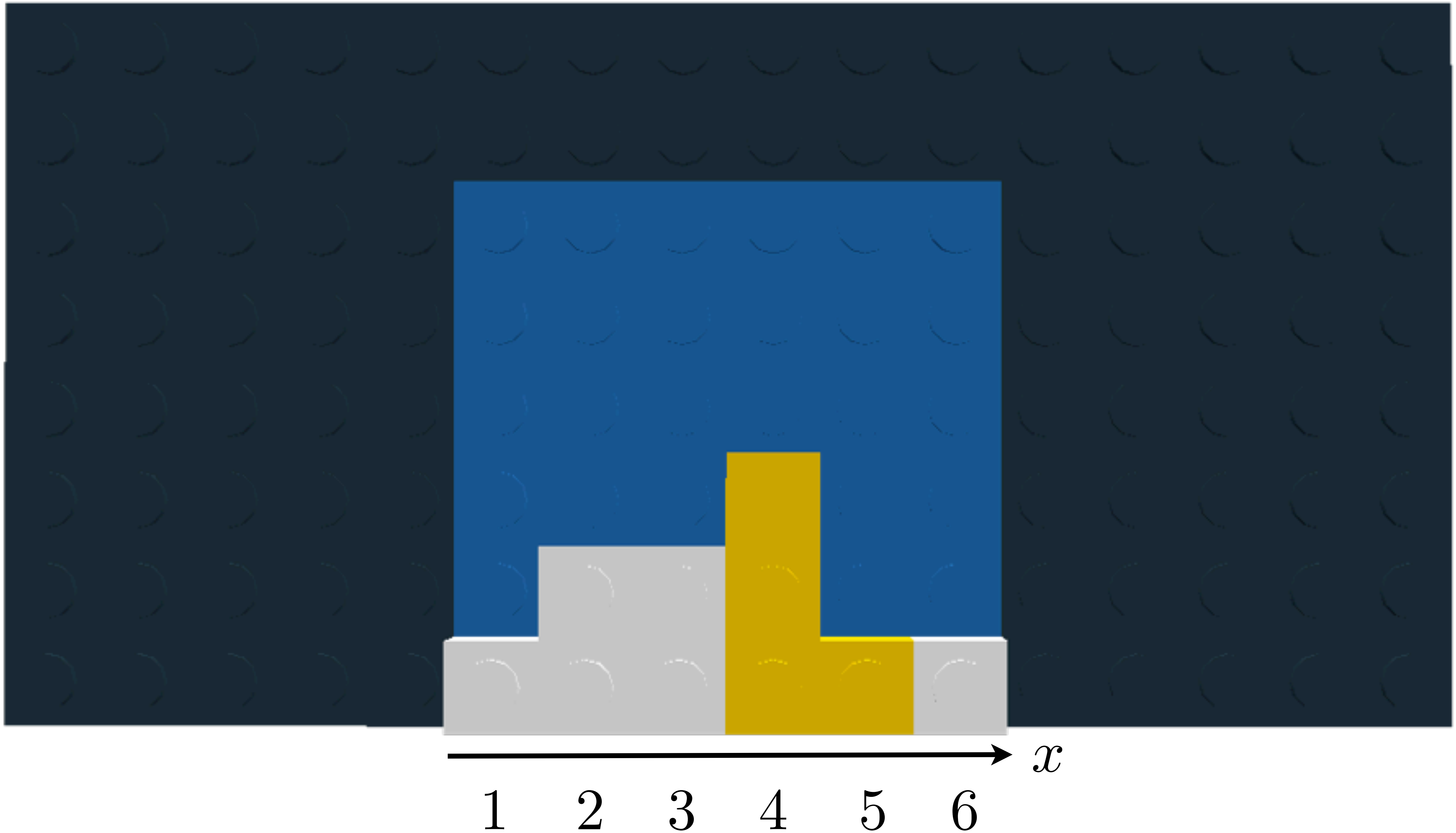
Hadley Wickham

1. Introduction and example
2. Multivariate calculus
3. Calculating probabilities
4. Expectation
5. The 2d cdf

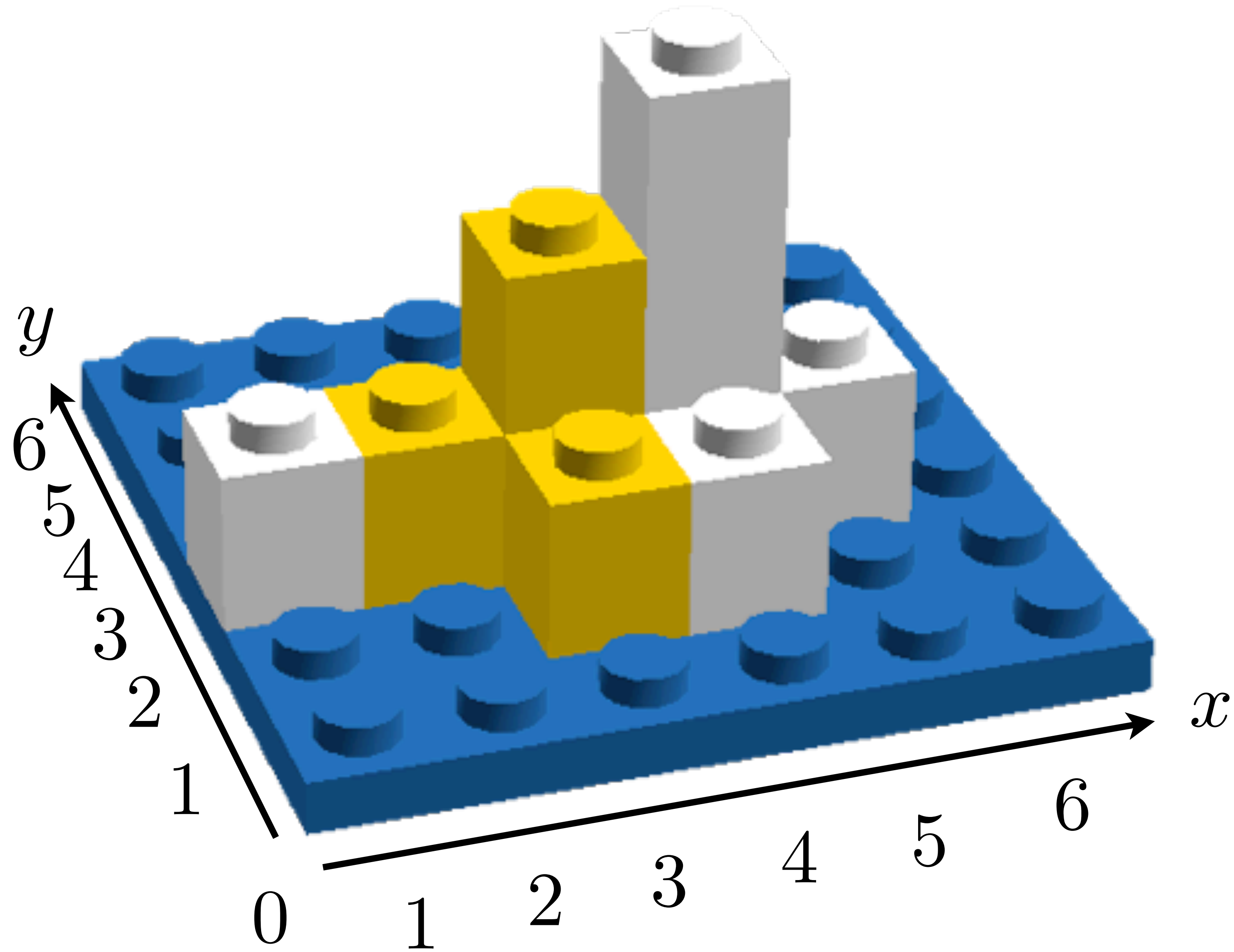
Introduction



$$\text{What is } P(Gold)?$$

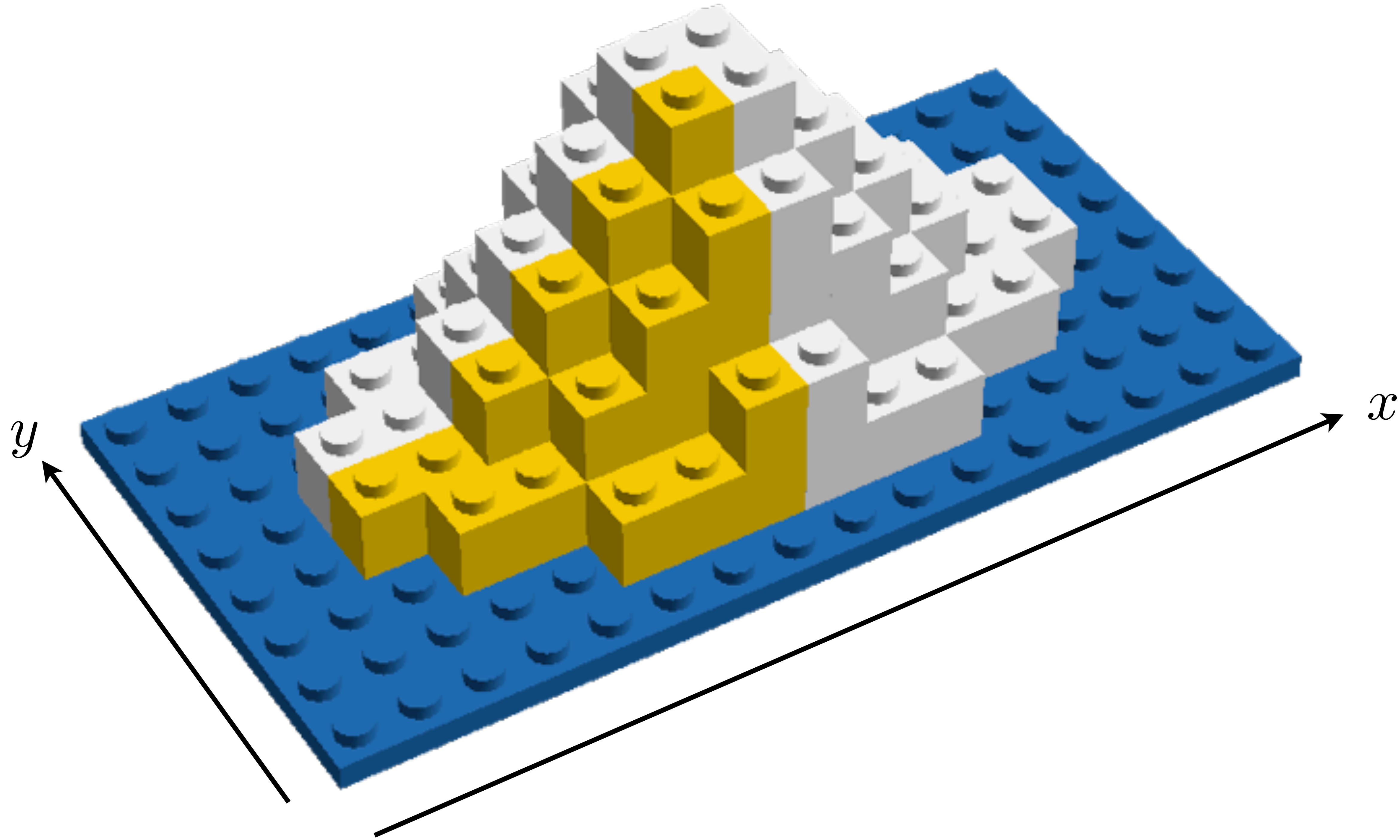


What is $P(Gold)$? $P(3 < X < 6)$?



What is $P(Gold)$?

$P(1 < X < 4,\ 1 < Y < 4)$?



Intuition: we're still just counting the possible events that fall into a region, but now our regions are 2D.

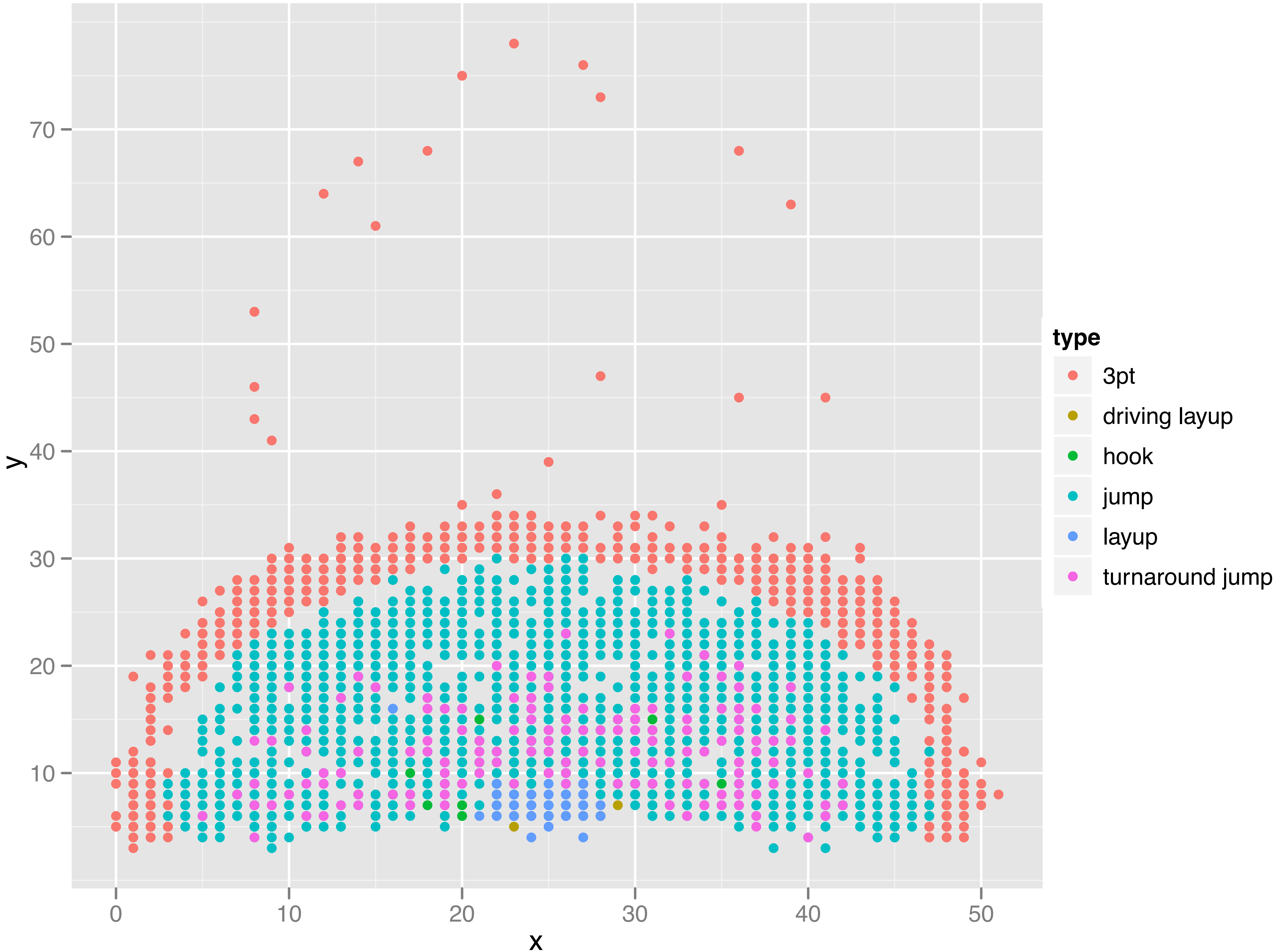
Bivariate rv

A **random experiment** where we measure **two things** (not just one). A vector instead of a single observation.

These variables could be both discrete, both continuous, or one continuous and one discrete. We will focus on both continuous: a **bivariate continuous random variable**.

Examples of Bivariate rv's

- Location: latitude, longitude
- SAT score: verbal, math
- Car mileage: city, highway
- Univariate measurements that are often looked at together (e.g, height and weight, income and crime, etc.)



70
60
50
40
30
20
10
y
0
10
20
30
40
50
x
type
3pt
driving layup
hook
jump
layup
turnaround jump

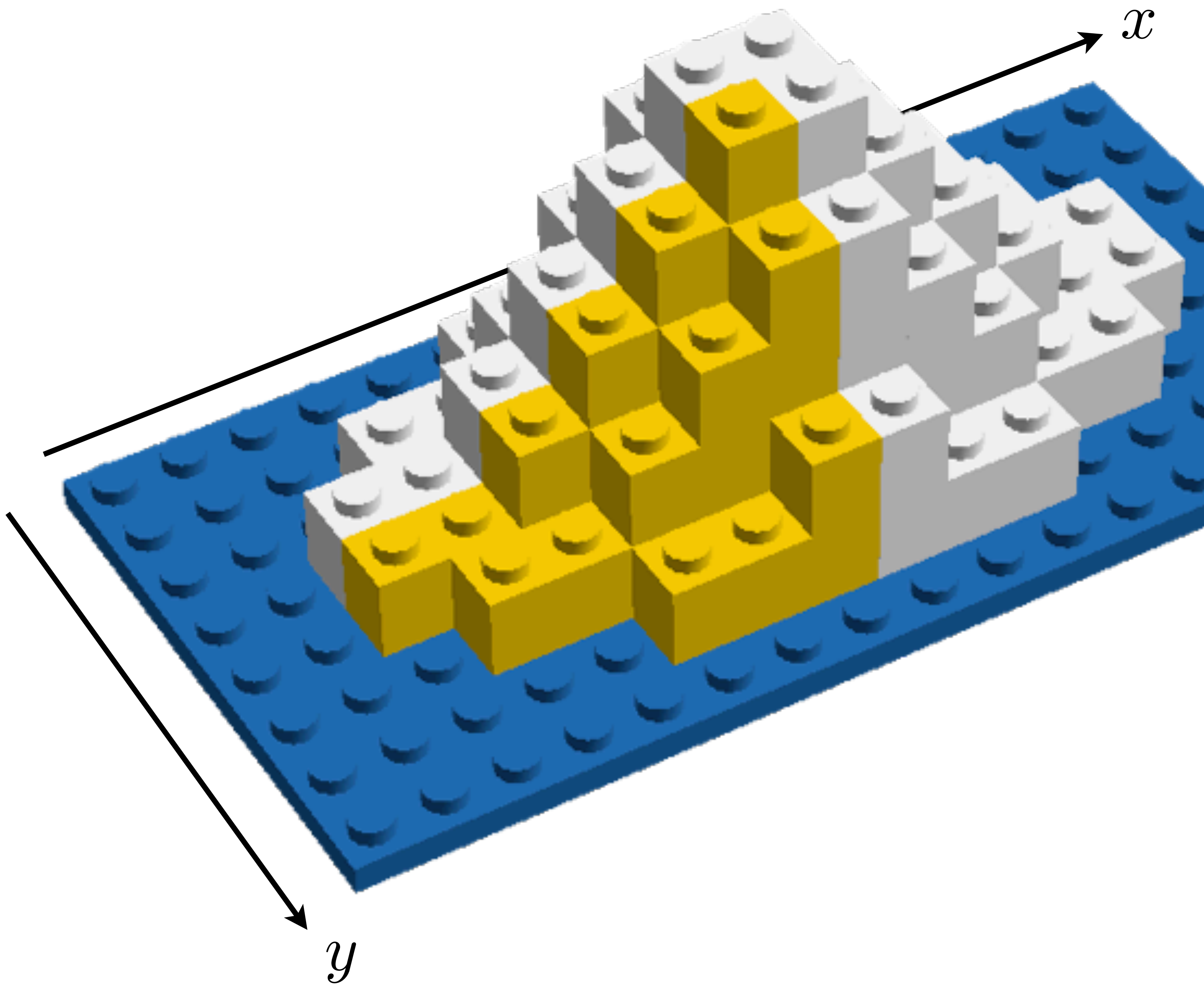
Sample space

Univariate continuous rv: sample space is an interval on the real line.

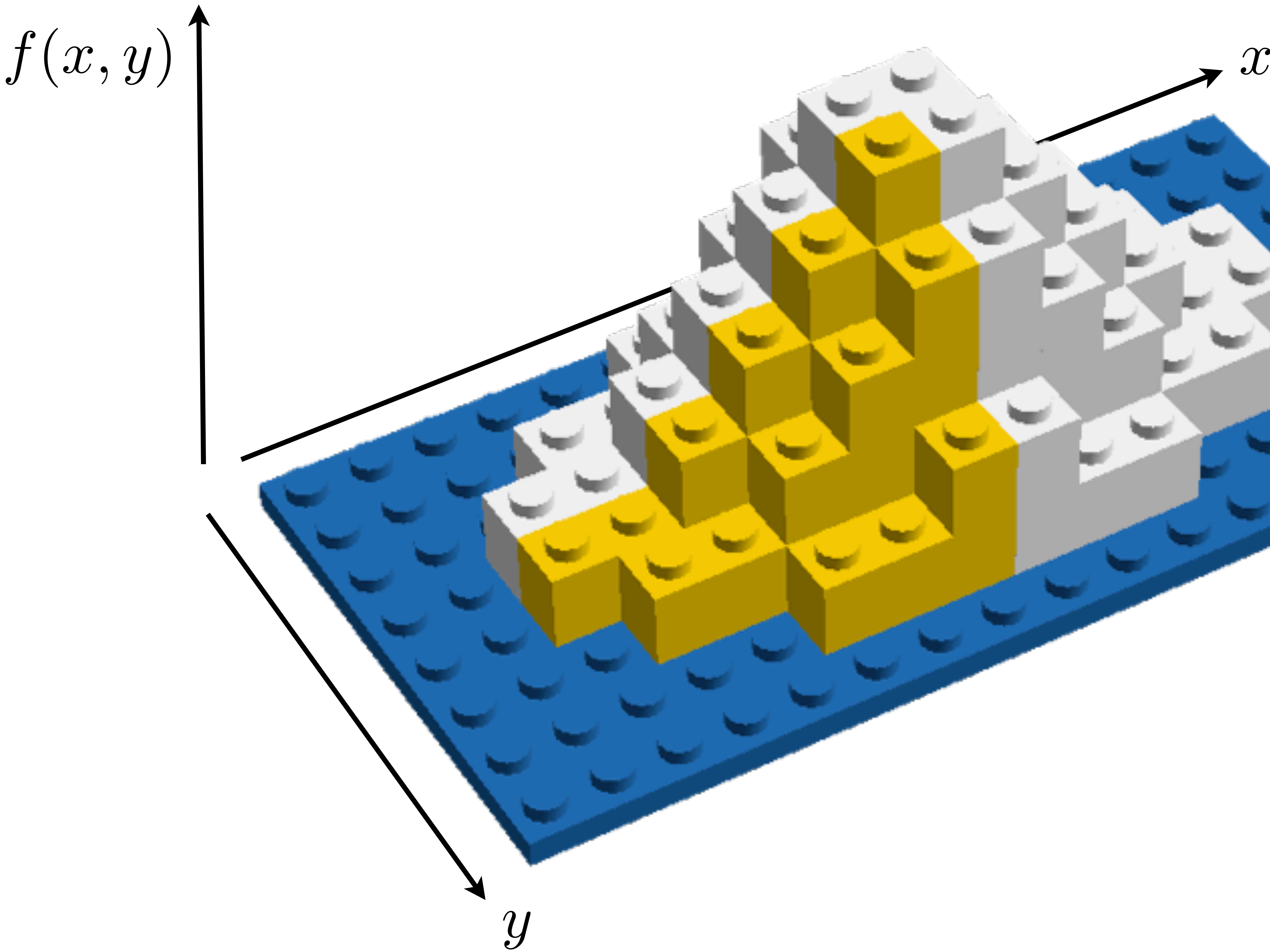
Bivariate continuous rv: sample space is region on the real plane.

$$S = \{(x, y) : f(x, y) > 0\}$$

What is the sample space? Which axis might show f(x,y)?



What is the sample space? Which axis might show f(x,y)?



Calculating probability

Intuition

In one dimension, probability was area under a curve.

In two dimensions, probability is volume under a surface.

Continuous case

$$P((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$$

$$P(x_1 < X < x_2, y_1 < Y < y_2) =$$

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x,y)\,dy\,dx$$

Discrete case

$$P((X,Y) \in A) = \sum_A \sum f(x,y)$$

$$P(X \in (i, i+1, ...n), y \in (j, j+1, ...m)) =$$

$$\sum_{x=i}^{n} \sum_{y=j}^{m} f(x,y)$$

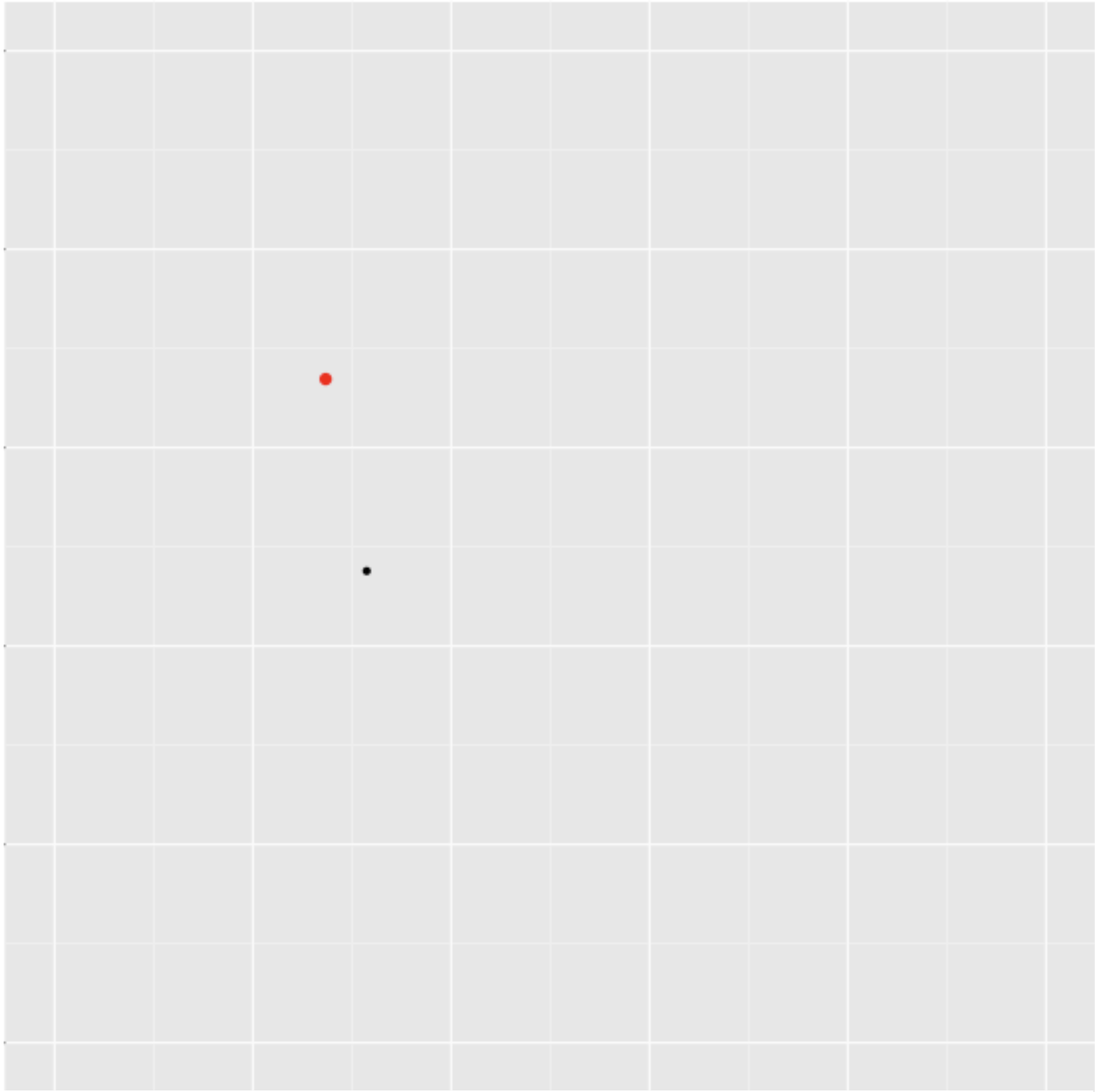
$$f(x,y) = \frac{1}{16} \qquad -2 < x, y < 2$$

What is:

- $P(X < 0)$?
- $P(X < 0 \text{ and } Y < 0)$?
- $P(Y > 1)$?
- $P(X > Y)$?
- $P(X^2 + Y^2 < 1)$

What would you call this distribution?

Draw diagrams and use your intuition



$$f(x, y) = c \quad a < x, y < b$$

Is this a pdf?

How could we work out c?

Your turn

Given what you know about univariate pdfs and pmfs, guess the conditions that a bivariate function must satisfy to be a bivariate pdf/pmf.

$$f(x, y) \geq 0 \quad \forall\, (x, y) \in \mathbb{R}^2$$

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

$$\iint_{\mathbb{R}^2} f(x, y)\ dy\ dx = 1$$

$$f(x,y) \geq 0 \quad \forall\,(x,y) \in S$$

$$\sum_{x,y \in S} \sum f(x,y) = 1$$

$$\sum_{x \in \mathbb{Z}} \sum_{y \in \mathbb{Z}} f(x,y) = 1$$

Your turn

$$f(x,y) = e^{-(x+y)} \quad x,y > 0$$

Is this a valid pdf?

Wolfram alpha

```
integrate_(x > 0) integate_(y > 0) e^(-x
- y) dx dy
```

Your turn

Convert these word problems into math problems, and then write the solution as an integral.

What's the probability that both X and Y are greater than 10?

What's the probability that X is bigger than 4 or Y is less than 3?

What's the probability that X is bigger than Y?

Wolfram alpha

```
int e^(-x-y) dx dy  x = 10 to inf, y =
10 to inf

1 - int e^(-x-y) dx dy  x = 0 to 4, y =
3 to inf

int e^(-x-y) dx dy  x = 0 to y, y = 0
to inf
```

CDF

What is the **cdf** going to look like?

What is the **cdf** going to look like?

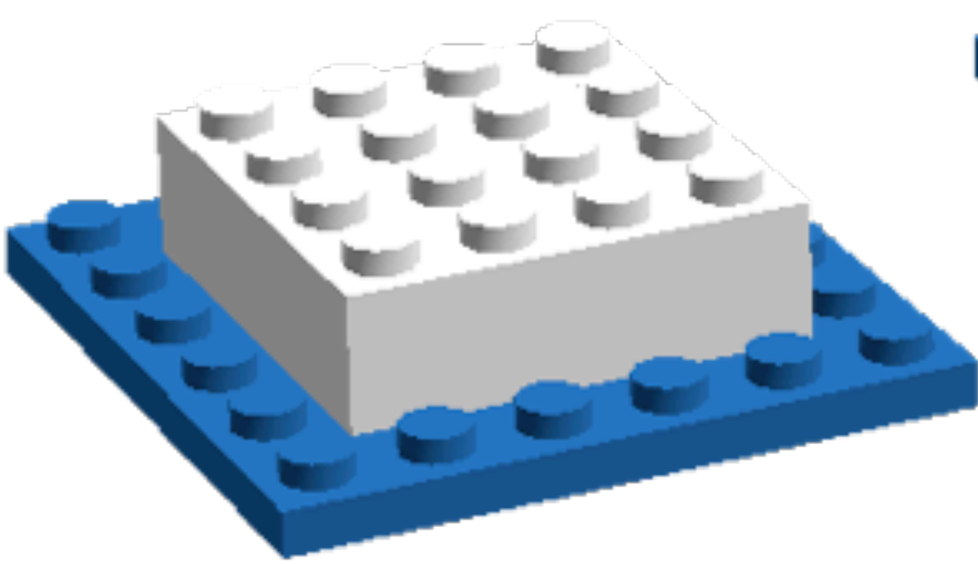
$$P(X < x, Y < y) =$$

What is the **cdf** going to look like?

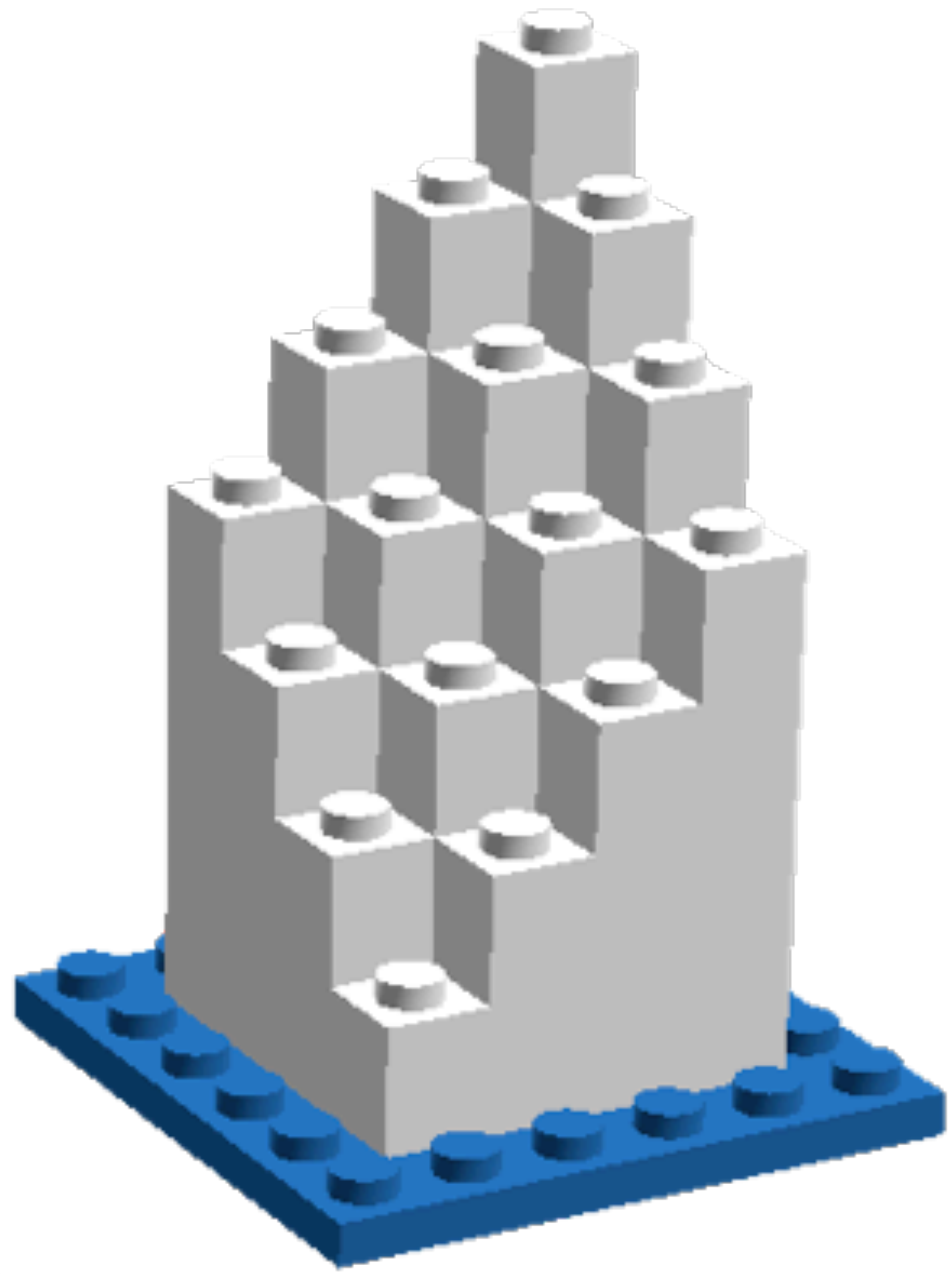
$$P(X < x, Y < y) =$$

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) dv du$$

Example: Bivariate Uniform



$f(x, y)$



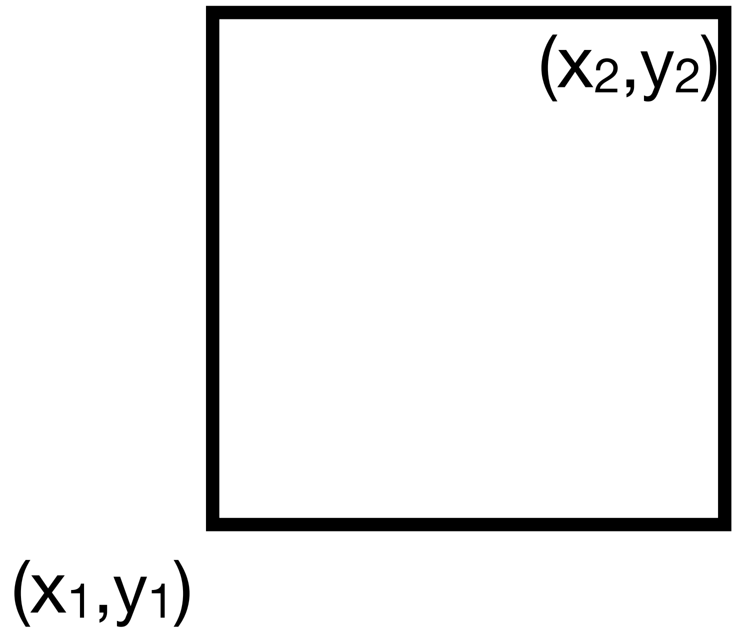
$F(x, y)$

Why is the cdf less useful in 2d?

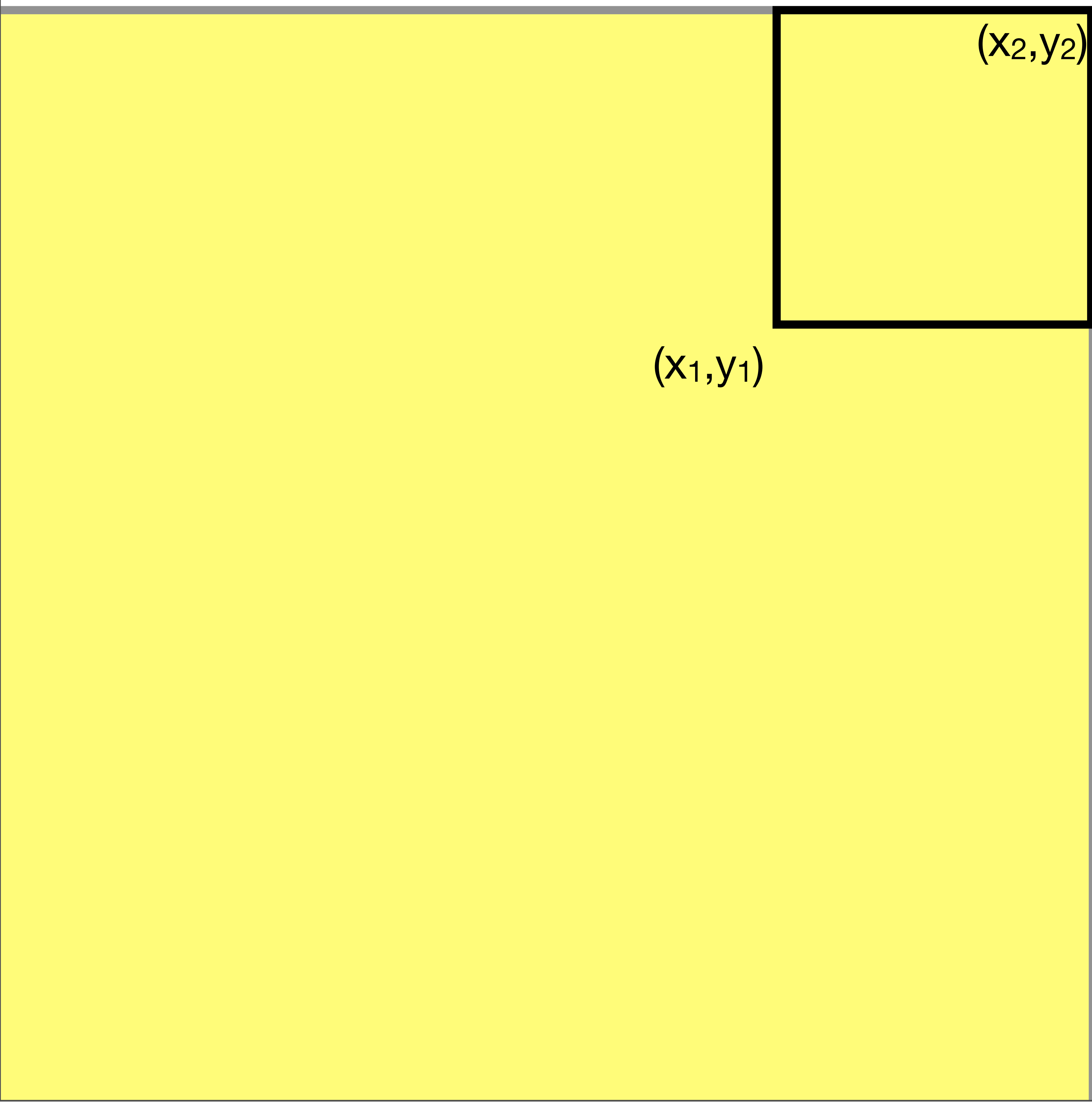
$$P(X^2 + Y^2 < 1)$$

$$P(x_1 < X < x_2, y_1 < Y < y_2)$$

$$P(x_1 < X < x_2, y_1 < Y < y_2) =$$

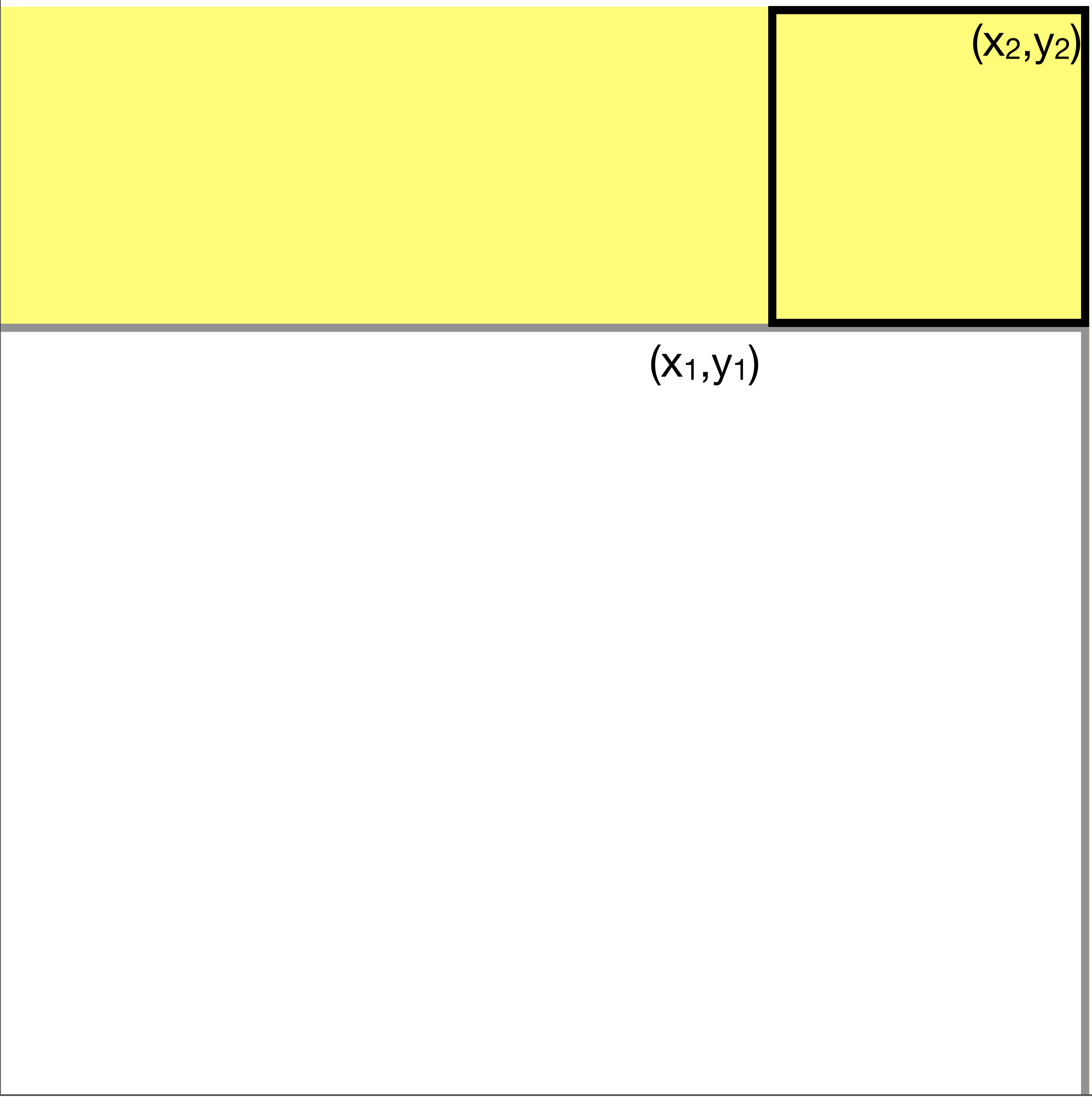


$$P(x_1 < X < x_2,\ y_1 < Y < y_2) =$$



$$F(x_2,\ y_2)$$

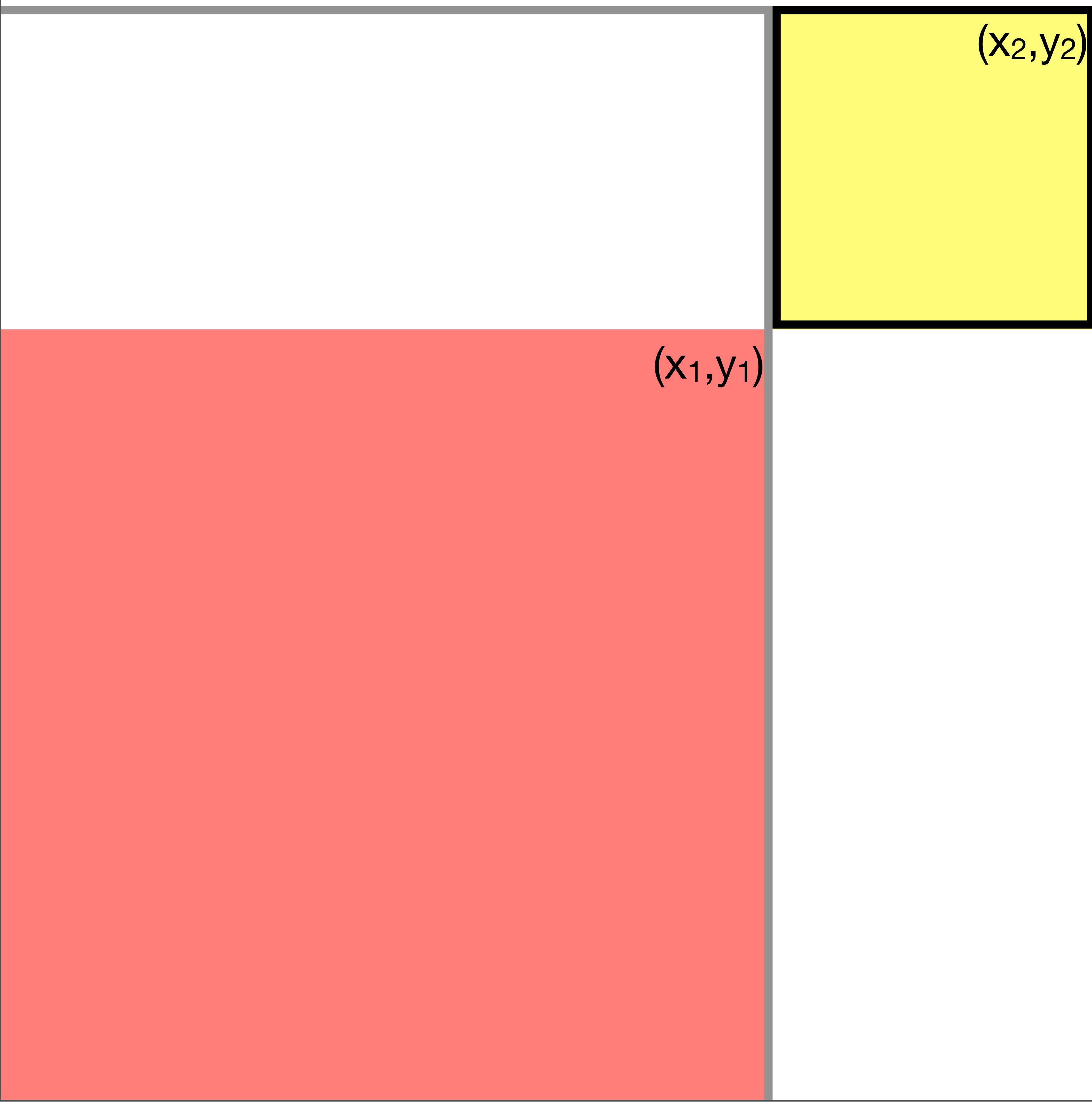
$$P(x_1 < X < x_2,\ y_1 < Y < y_2) =$$



$F(x_2,\ y_2)$

$- F(x_1,\ y_2)$

$$P(x_1 < X < x_2,\ y_1 < Y < y_2) =$$

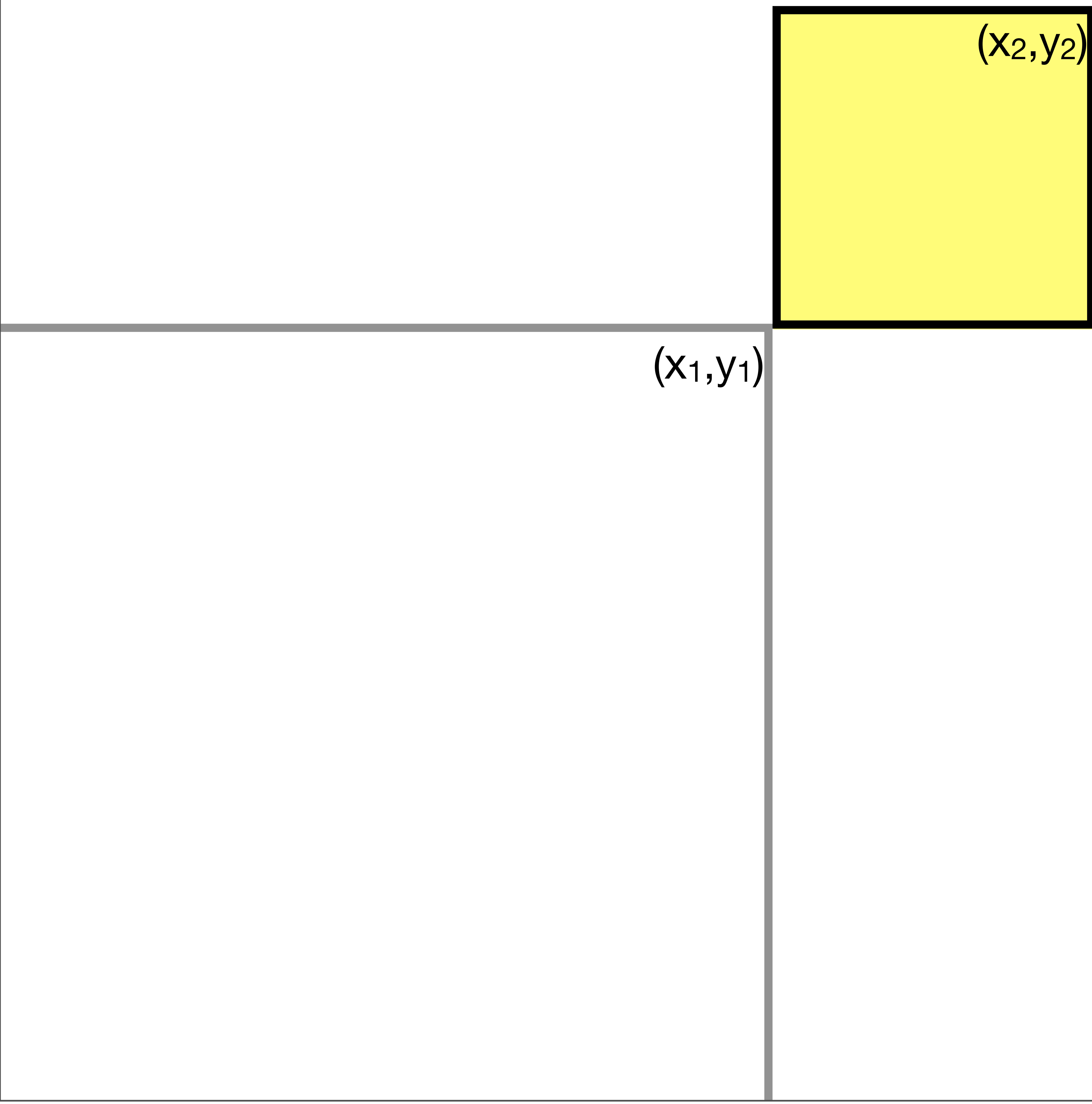


$$F(x_2,\ y_2)$$

$$-\ F(x_1,\ y_2)$$

$$\mathbf{-\ F(x_2,\ y_1)}$$

$$P(x_1 < X < x_2,\ y_1 < Y < y_2) =$$



$$F(x_2,\ y_2)$$

$$-\ F(x_1,\ y_2)$$

$$-\ F(x_2,\ y_1)$$

$$\mathbf{+\ F(x_1,\ y_1)}$$

Your turn

$F(x, y) = cxy(x + y) \quad 0 < x, y < 2$

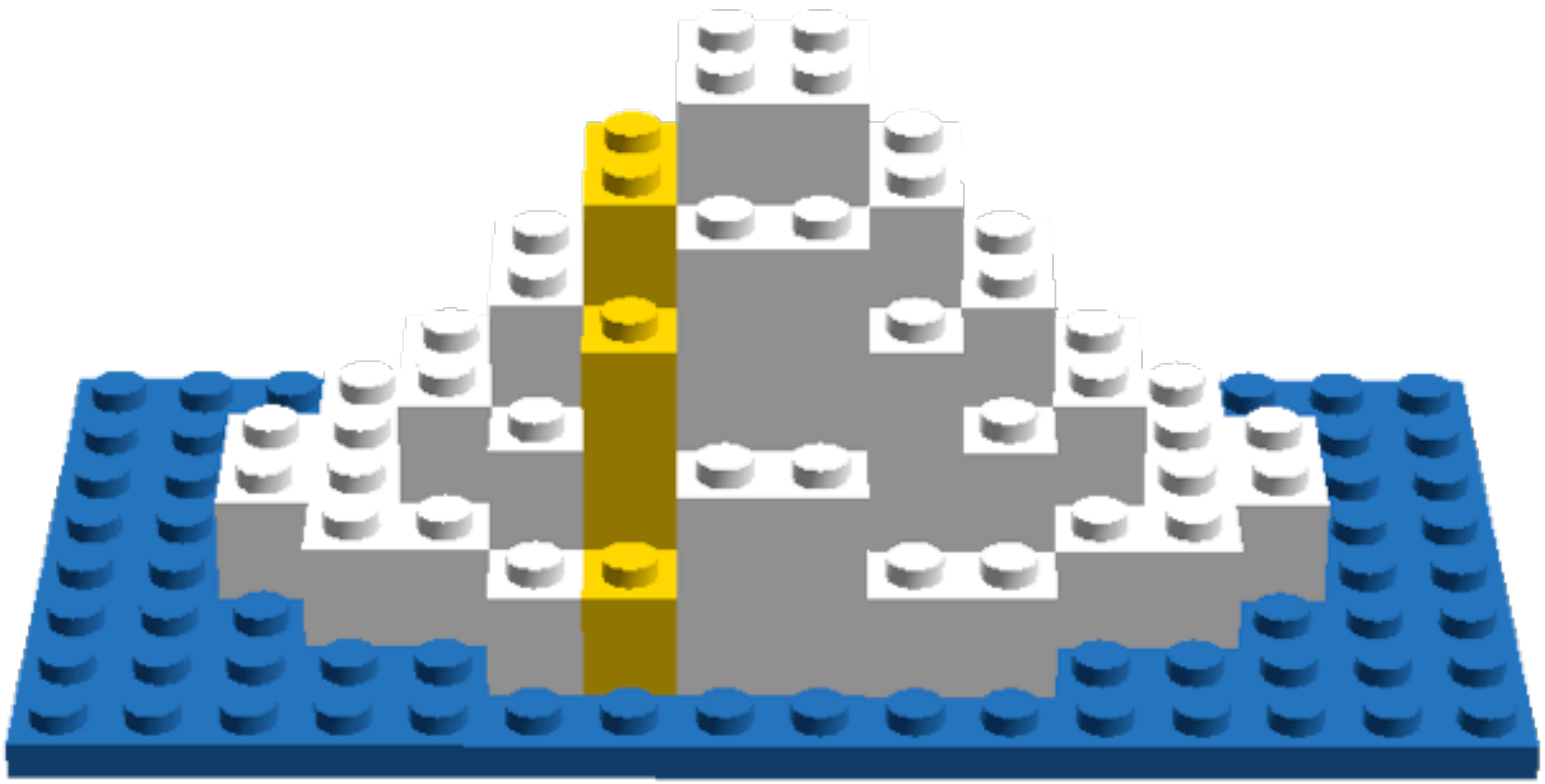
What is c?

What is f(x, y)?

CDF → PDF

Need to differentiate once for each variable.

Next time



What is $P(X = 7)$? How could we rearrange the above to just get the *pdf* of X?