

ML Basic

NumPy

Начнем в 20:01 мск

otus.ru



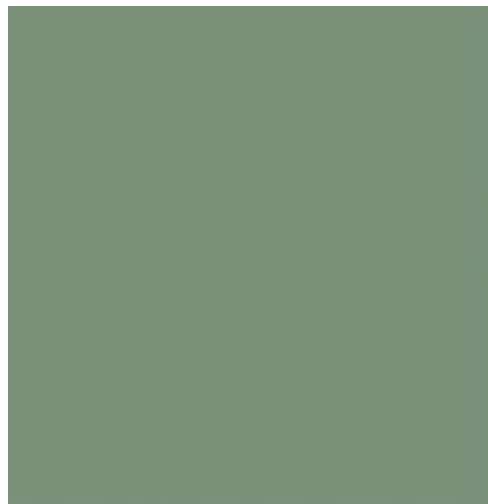
Какие
животные тут
зашифрованы?



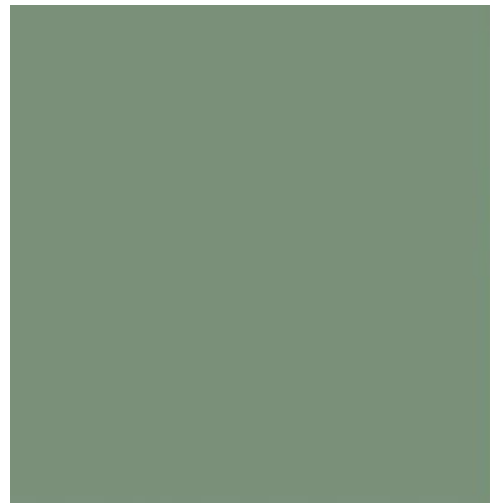
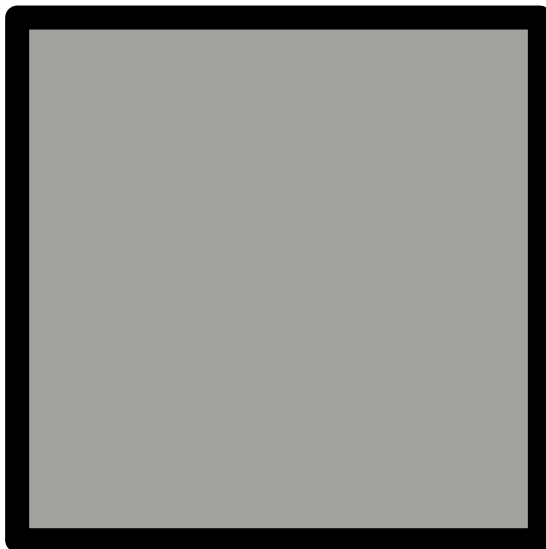


Если запись не включена,
КРИЧИТЕ 😱!

Животные



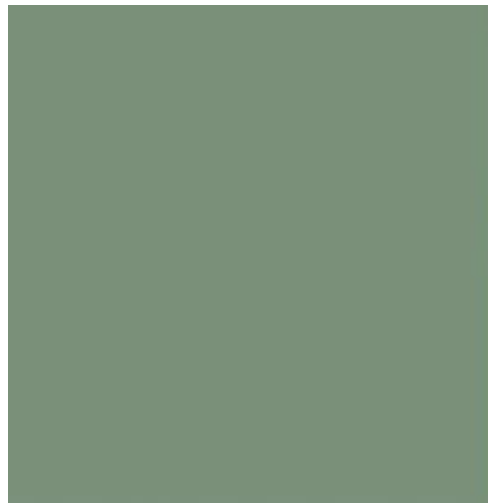
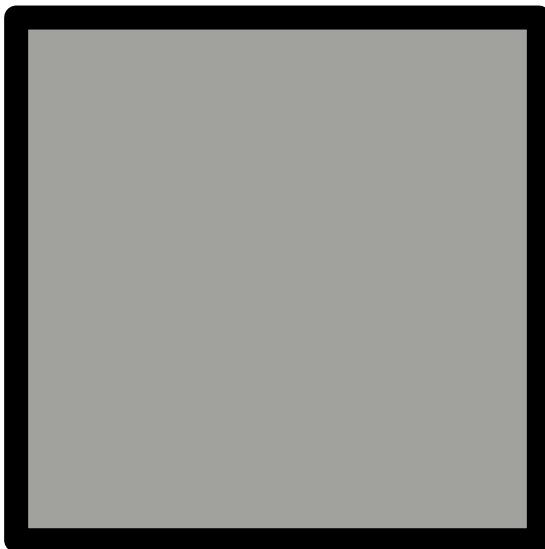
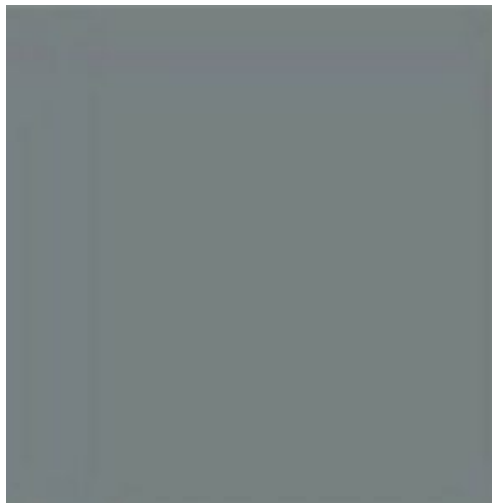
Животные



Животные



Испуганная
мышь

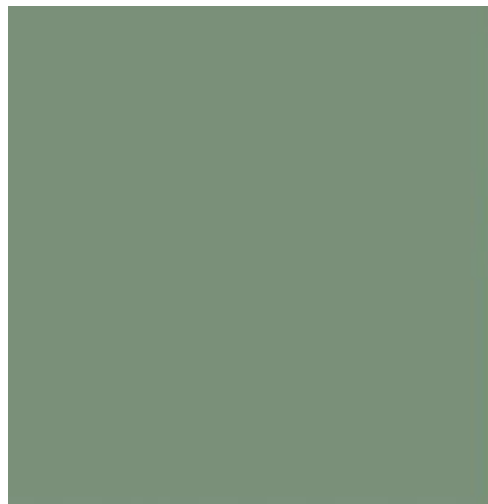
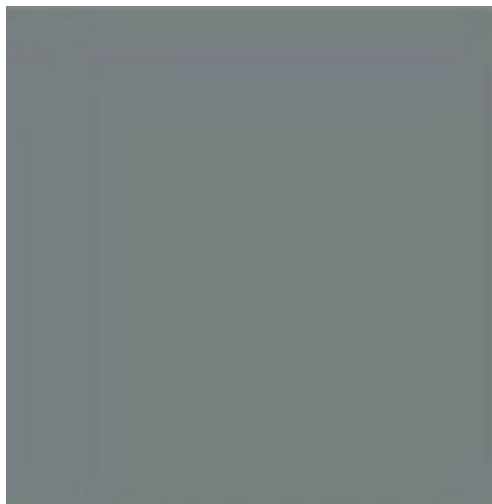




Названия цветов зафиксированы в различных системах и каталогах, среди которых:

- **Каталог RAL.** Разработан немецким институтом гарантий качества и сертификации в 1927 году. Часто применяемые цвета, каждому присвоен уникальный номер.
- **Система NCS** (Natural Color System). Разработана в 1979 году в Швеции. Каждый цвет закодирован по яркости, темноте и оттенку. Всего 2000 оттенков.
- **Система PMS** (Pantone Matching System). Признана во всём мире и используется в разных сферах: от полиграфии до дизайна интерьера и моды.

Животные



Животные



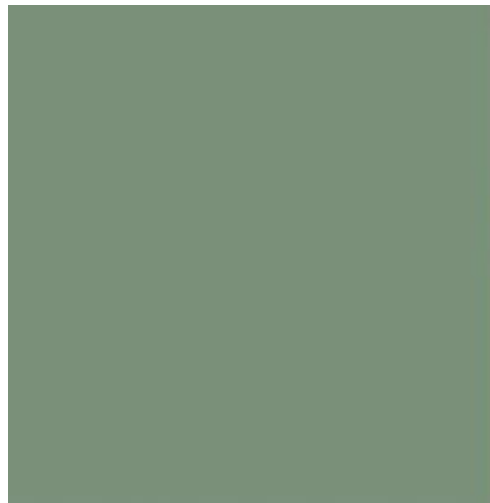
Цвет паука,
замышляющего
злодейство

Животные



Цвет паука,
замышляющего
злодейство

Цвет испуганной
мыши



Животные



Цвет паука,
замышляющего
злодейство

Цвет испуганной
мыши

Лягушка в
обмороке

[Дальше](#) ➞

Цель



- Научиться кодировать разные объекты “хорошо”

Цель



- Научиться кодировать разные объекты “хорошо”

Цель



- Научиться кодировать разные объекты “хорошо”
 - Числом
 - Есть возможность сравнения разных объектов
 - Разница между объектами интерпретируема

Цель



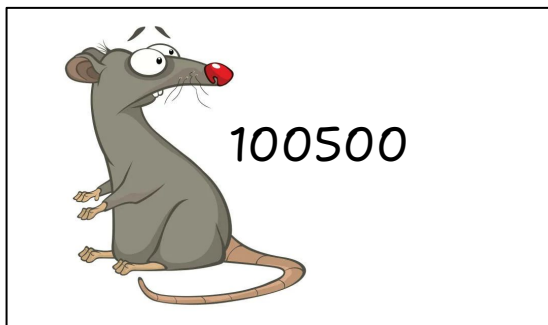
- Научиться кодировать разные объекты “хорошо”
 - Числом
 - Есть возможность сравнения разных объектов
 - Разница между объектами интерпретируема



Цель



- Научиться кодировать разные объекты “хорошо”
 - Числом
 - Есть возможность сравнения разных объектов
 - Разница между объектами интерпретируема

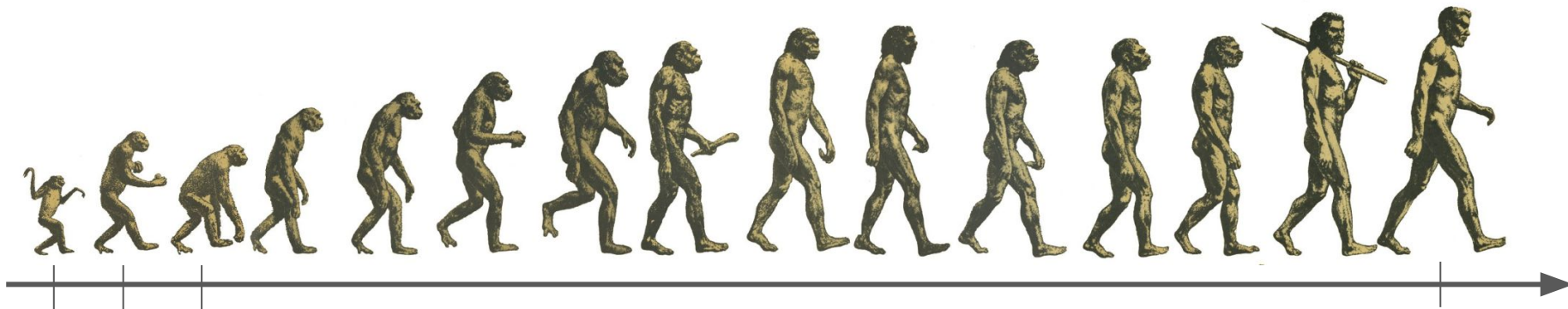


Цель



- Научиться кодировать разные объекты “хорошо”

- Числом
- Есть возможность сравнения разных объектов
- Разница между объектами интерпретируема



Как компьютер работает со словами?

Слово → число

Afghanistan
Albania
Algeria
Andorra
Angola
Antigua and Barbuda
Argentina
Armenia
Australia
Austria
Azerbaijan
Bahamas
Bahrain
Bangladesh
Barbados
Belarus

Как компьютер работает со словами?

Слово → число

1	Afghanistan
2	Albania
3	Algeria
4	Andorra
5	Angola
6	Antigua and Barbuda
	Argentina
	Armenia
	Australia
	Austria
	Azerbaijan
	Bahamas
	Bahrain
	Bangladesh
	Barbados
	Belarus

Как компьютер работает со словами?

Слово → число

1	Afghanistan
2	Albania
3	Algeria
4	Andorra
5	Angola
6	Antigua and Barbuda
	Argentina
	Armenia
	Australia
	Austria
	Azerbaijan
	Bahamas
	Bahrain
	Bangladesh
	Barbados
	Belarus

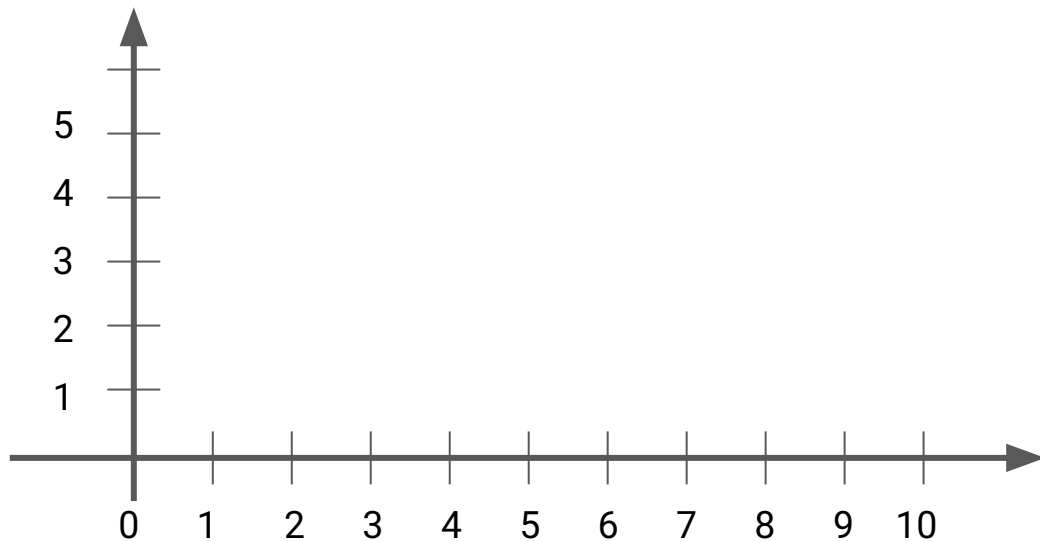
Как компьютер работает со словами?

Слово → число



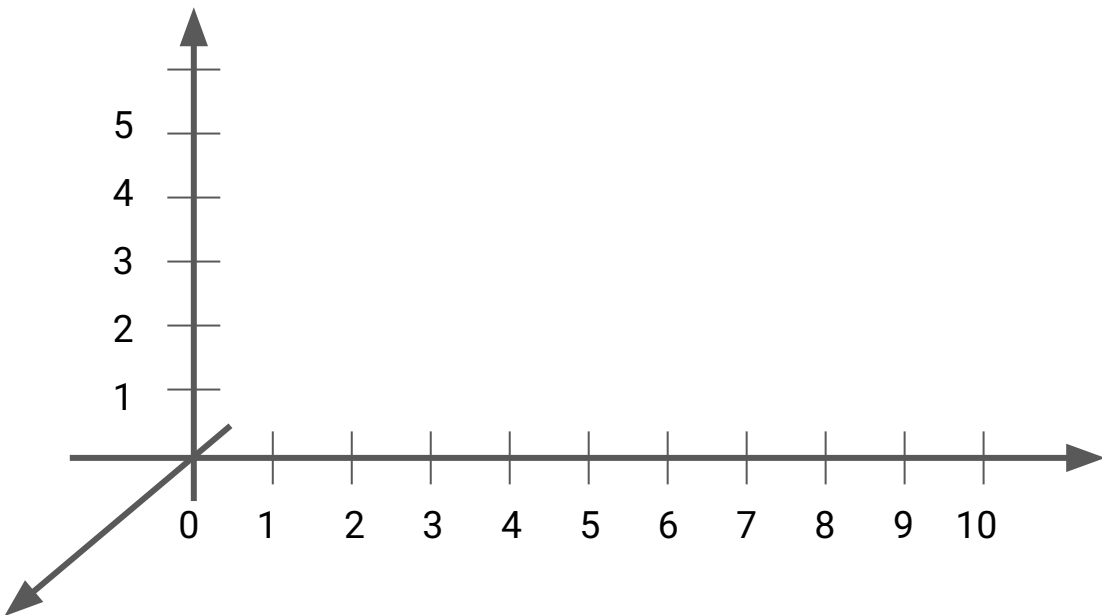
1	Afghanistan
2	Albania
3	Algeria
4	Andorra
5	Angola
6	Antigua and Barbuda
	Argentina
	Armenia
	Australia
	Austria
	Azerbaijan
	Bahamas
	Bahrain
	Bangladesh
	Barbados
	Belarus

Как компьютер работает со словами?



1	Afghanistan
2	Albania
3	Algeria
4	Andorra
5	Angola
6	Antigua and Barbuda
	Argentina
	Armenia
	Australia
	Austria
	Azerbaijan
	Bahamas
	Bahrain
	Bangladesh
	Barbados
	Belarus

Как компьютер работает со словами?



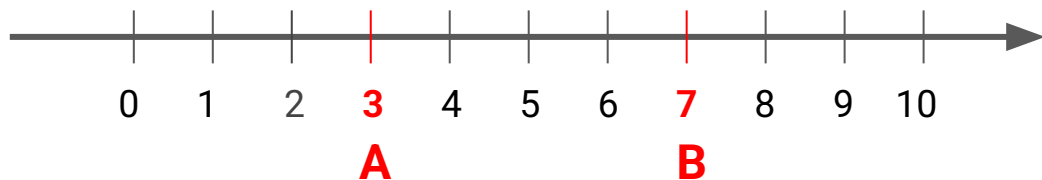
- | | |
|---|---------------------|
| 1 | Afghanistan |
| 2 | Albania |
| 3 | Algeria |
| 4 | Andorra |
| 5 | Angola |
| 6 | Antigua and Barbuda |
| | Argentina |
| | Armenia |
| | Australia |
| | Austria |
| | Azerbaijan |
| | Bahamas |
| | Bahrain |
| | Bangladesh |
| | Barbados |
| | Belarus |

Про расстояния



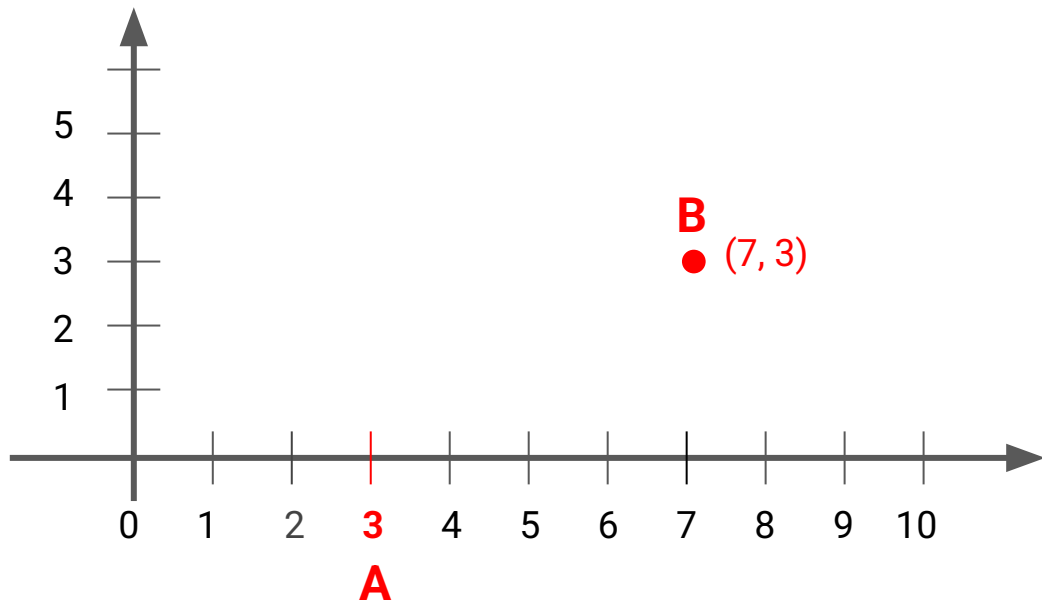
Найдите расстояние между точками А и В.

Про расстояния



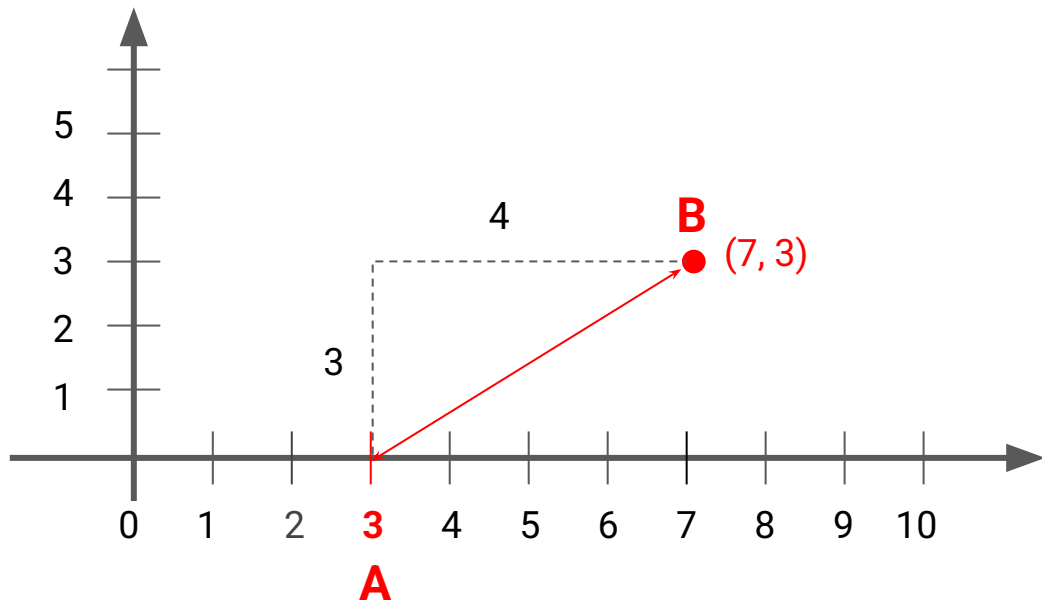
Найдите расстояние между точками A и B.

Про расстояния



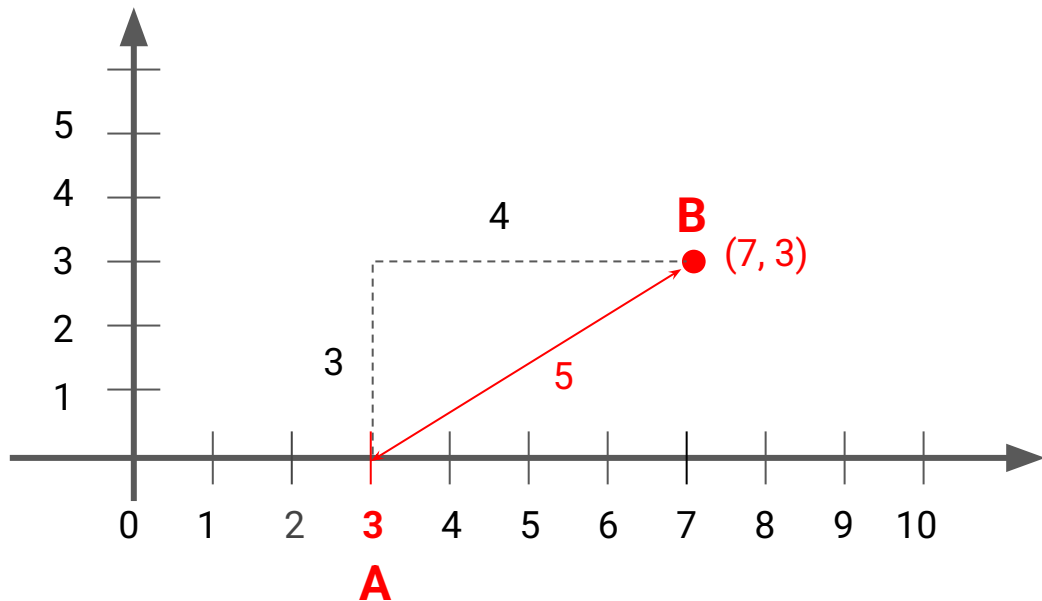
Найдите расстояние между точками A и B.

Про расстояния



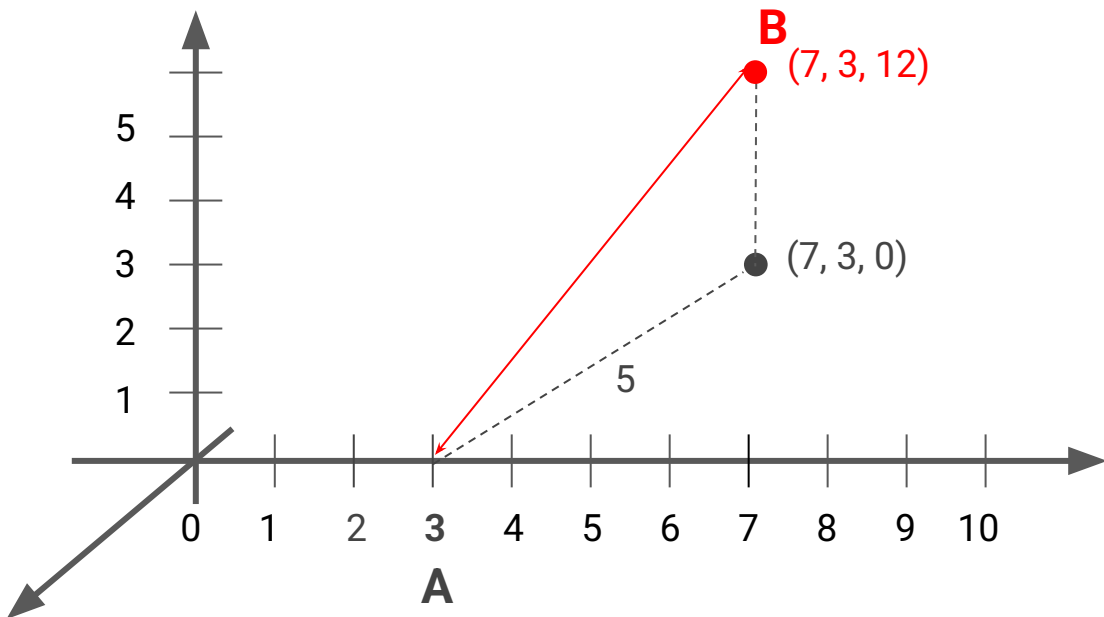
Найдите расстояние между точками A и B.

Про расстояния



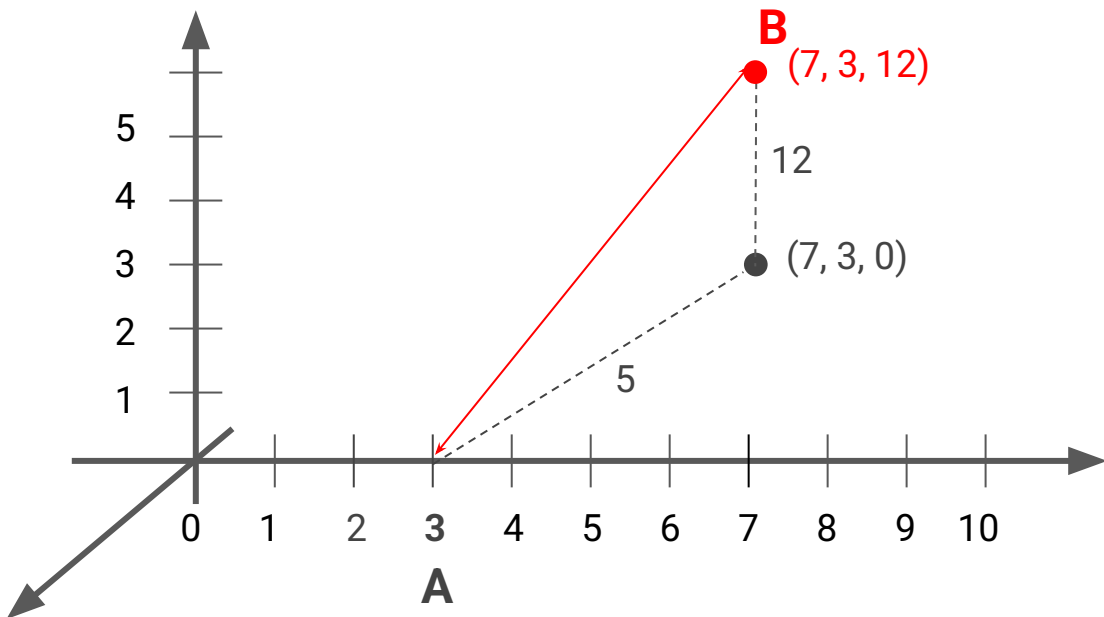
Найдите расстояние между точками A и B.

Про расстояния



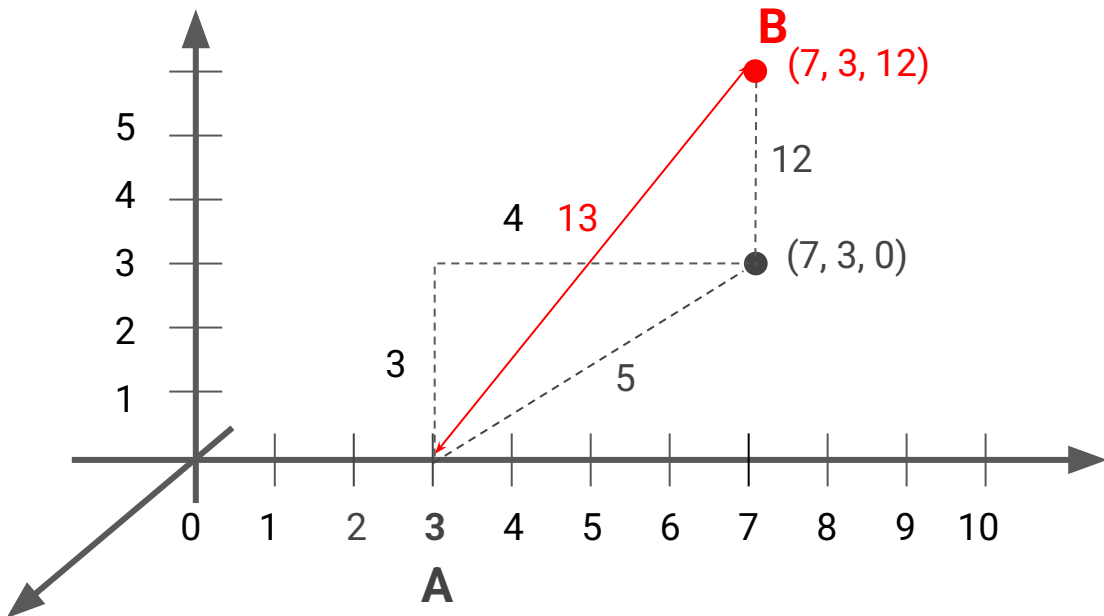
Найдите расстояние между точками A и B .

Про расстояния



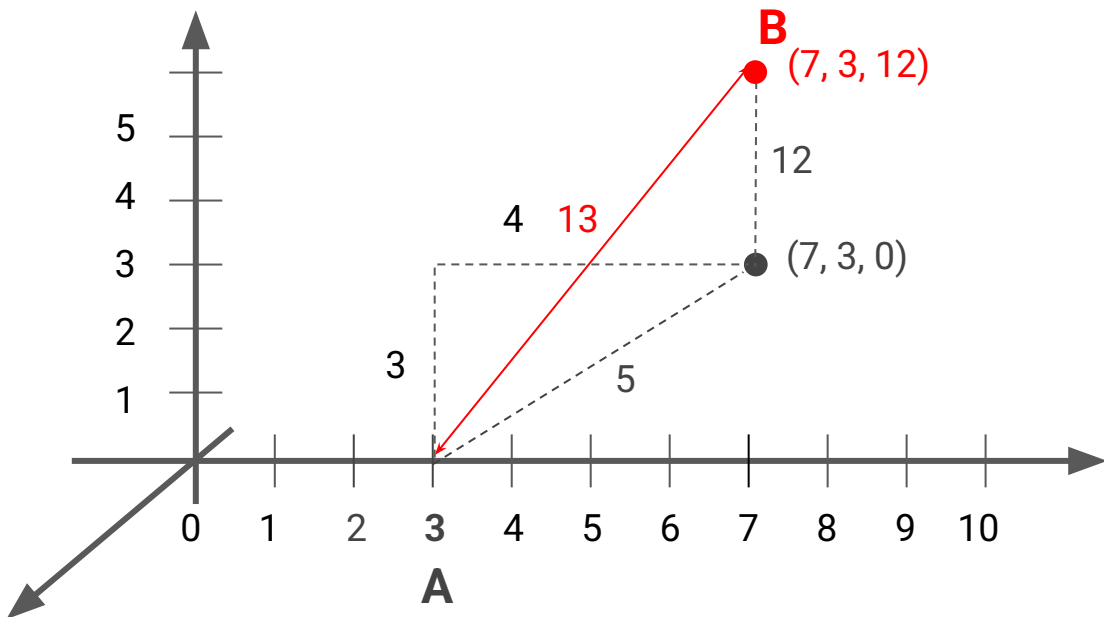
Найдите расстояние между точками A и B.

Про расстояния



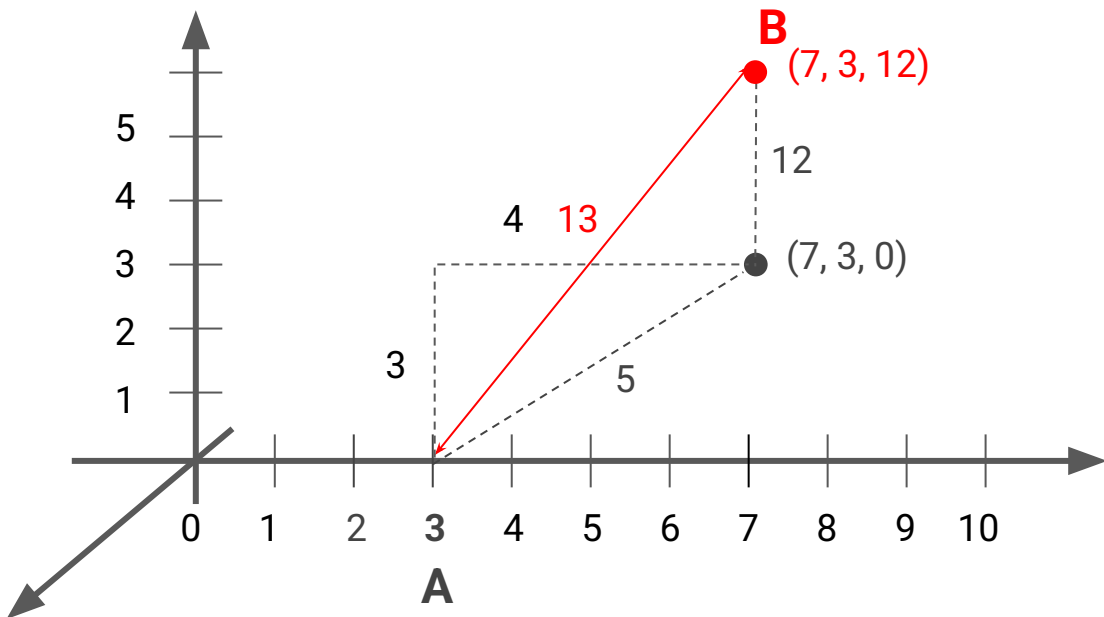
Найдите расстояние между точками A и B.

Итого



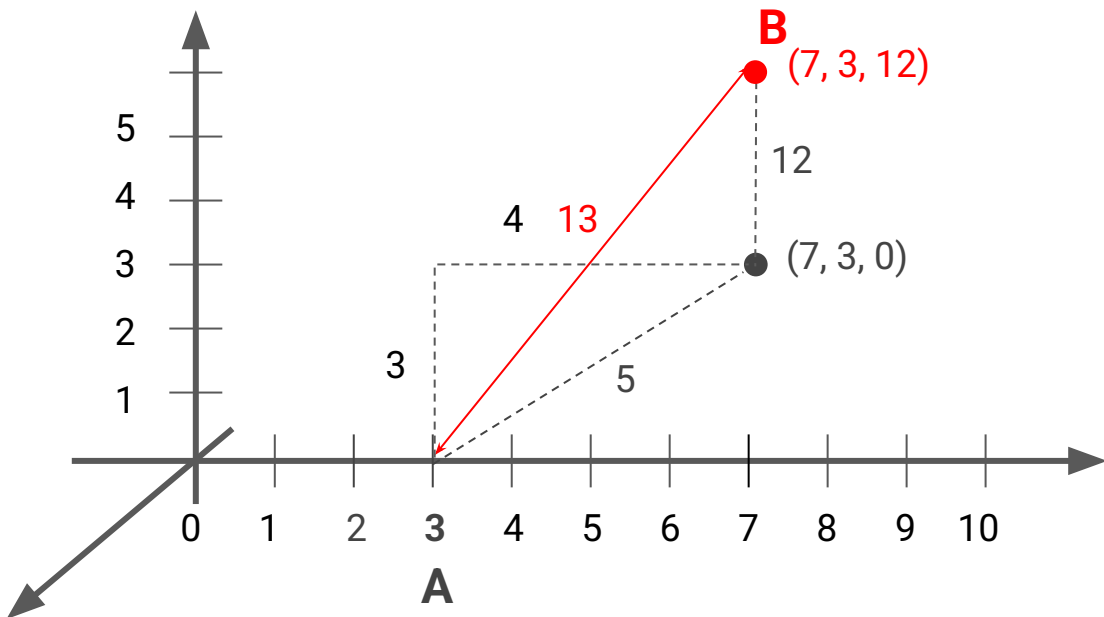
n	A	B	$\rho(A, B)$
1	(3,)	(7,)	4
2			
3			
4			
...			

Итого



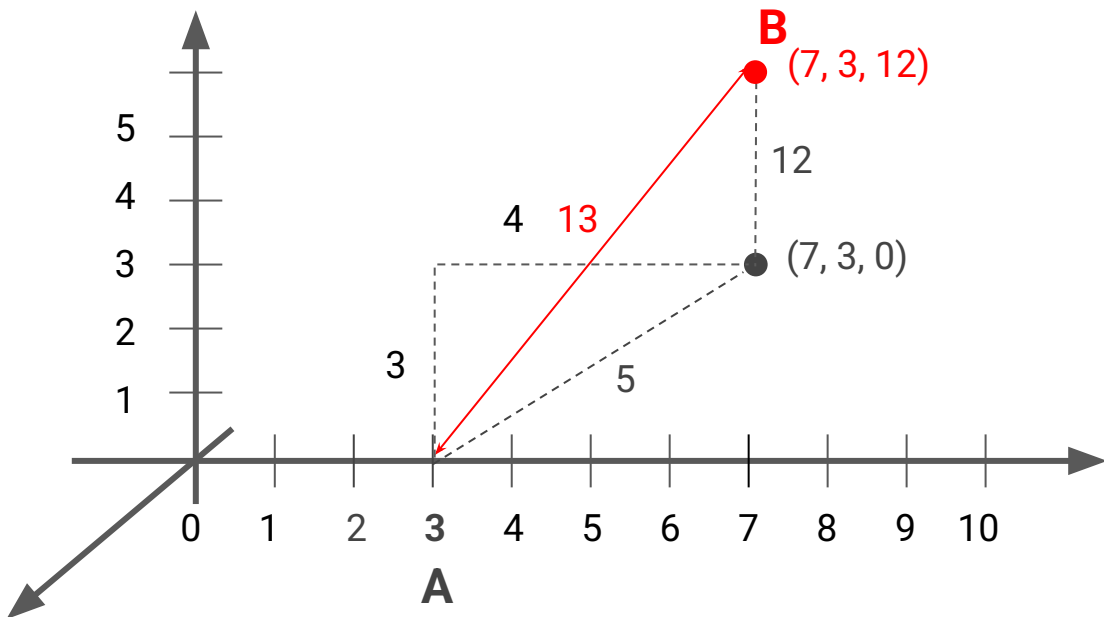
n	A	B	$\rho(A, B)$
1	(3,)	(7,)	4
2	(3, 0)	(7, 3)	5
3			
4			
...			

Итого



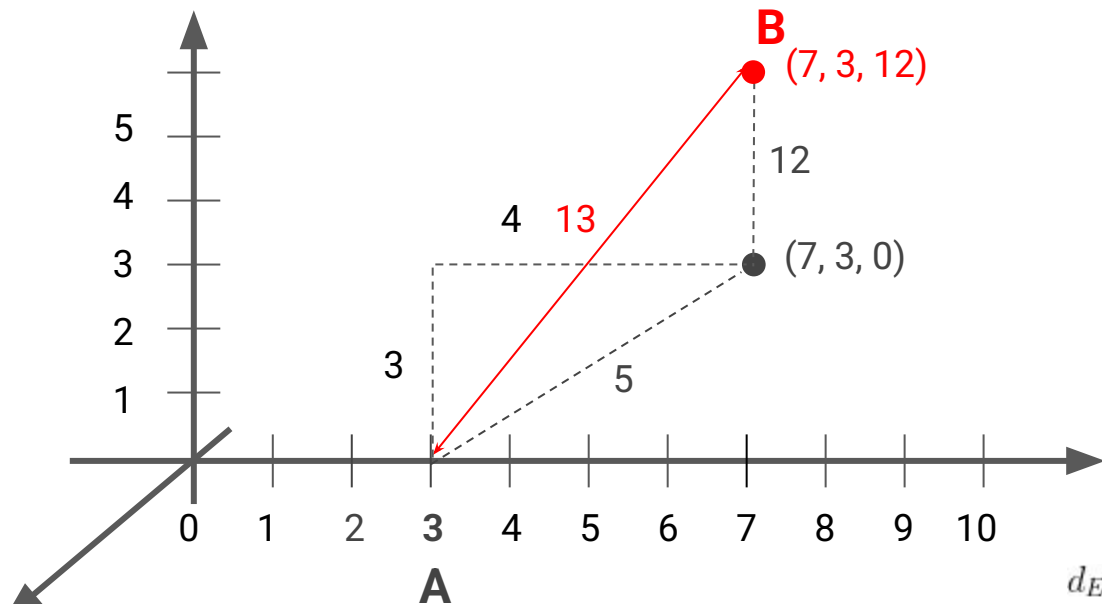
n	A	B	$\rho(A, B)$
1	(3,)	(7,)	4
2	(3, 0)	(7, 3)	5
3	(3, 0, 0)	(7, 3, 12)	13
4			
...			

Итого



n	A	B	$\rho(A, B)$
1	(3,)	(7,)	4
2	(3, 0)	(7, 3)	5
3	(3, 0, 0)	(7, 3, 12)	13
4	(3, 0, 0, 0)	(7, 3, 12, 84)	
...			

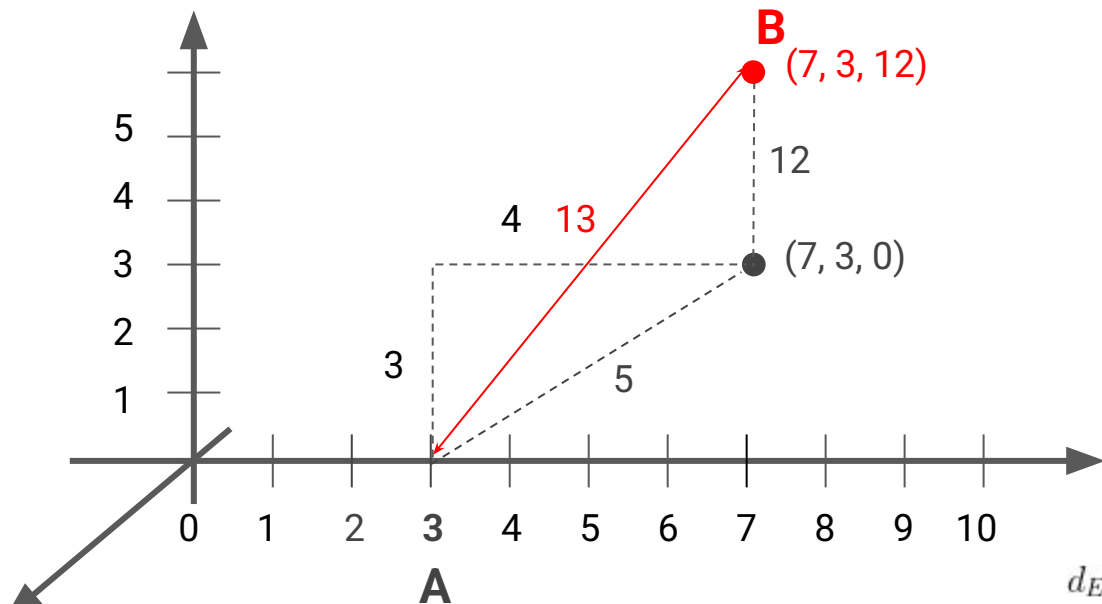
Итого



n	A	B	$\rho(A, B)$
1	(3,)	(7,)	4
2	(3, 0)	(7, 3)	5
3	(3, 0, 0)	(7, 3, 12)	13
4	(3, 0, 0, 0)	(7, 3, 12, 84)	
...			

$$\begin{aligned}
 d_{Eucl}(p, q) &= \\
 &= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \\
 &= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}
 \end{aligned}$$

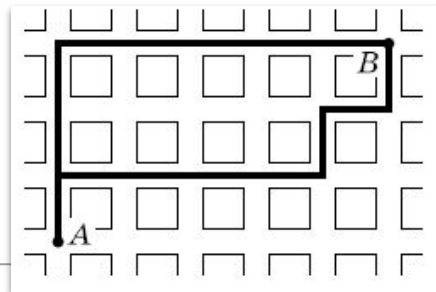
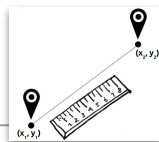
Итого



n	A	B	$\rho(A, B)$
1	(3,)	(7,)	4
2	(3, 0)	(7, 3)	5
3	(3, 0, 0)	(7, 3, 12)	13
4	(3, 0, 0, 0)	(7, 3, 12, 84)	85
...			

$$\begin{aligned}
 d_{Eucl}(p, q) &= \\
 &= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \\
 &= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}
 \end{aligned}$$

Аксиомы метрического пространства

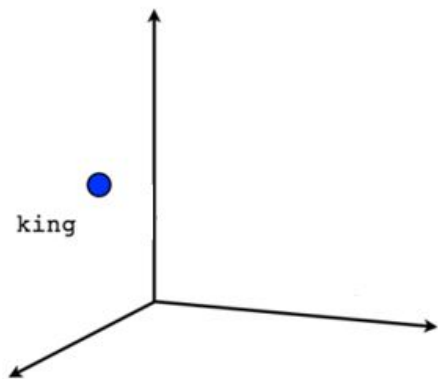


$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

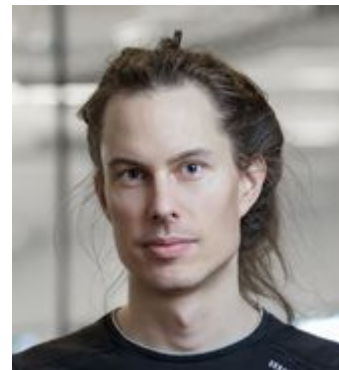
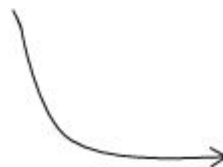
$$|x_2 - x_1| + |y_2 - y_1|$$

1	$\varrho(X, Y) \geq 0$	+	+
2	$\varrho(X, Y) = 0 \Leftrightarrow X = Y$	+	+
3	$\varrho(X, Y) = \varrho(Y, X)$	+	+
4	$\varrho(X, Y) \leq \varrho(X, Z) + \varrho(Z, Y)$	+	+

word2vec

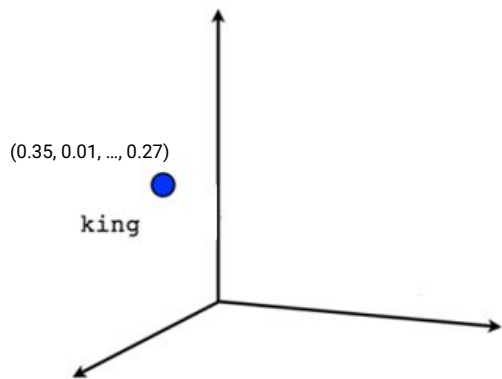


Tomas Mikolov
2013

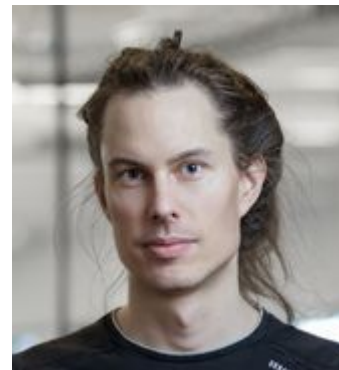
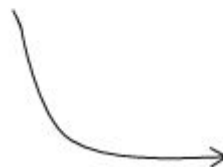


Подобрать такие вектора, чтобы чем чаще слова встречаются в одном контексте, тем они ближе.

word2vec

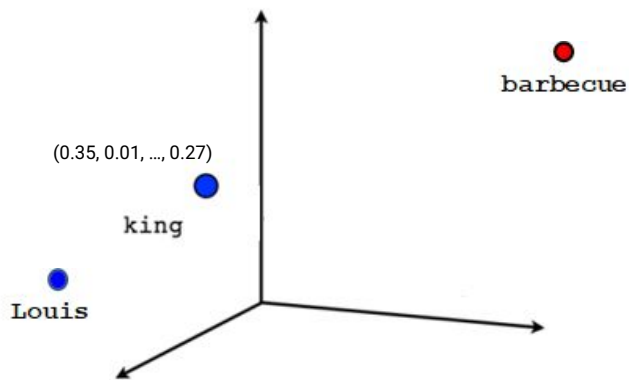


Tomas Mikolov
2013

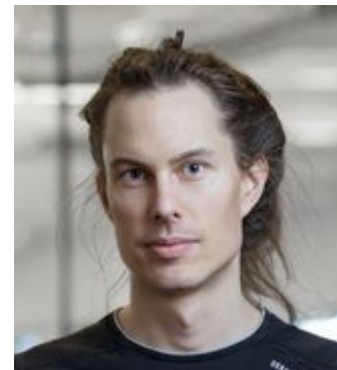
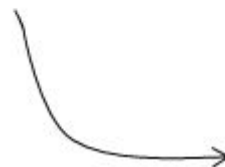


Подобрать такие вектора, чтобы чем чаще слова встречаются в одном контексте, тем они ближе.

word2vec

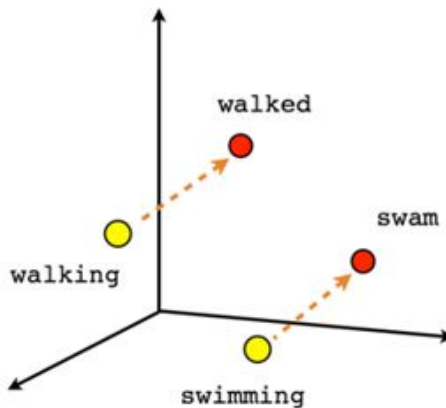
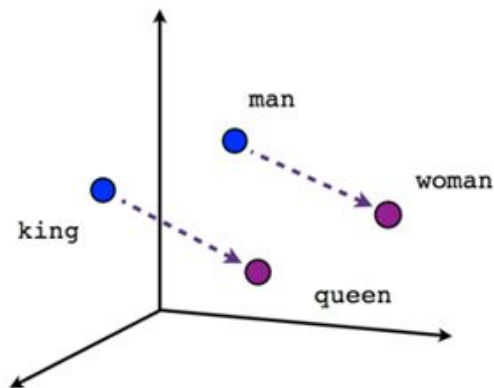


Tomas Mikolov
2013

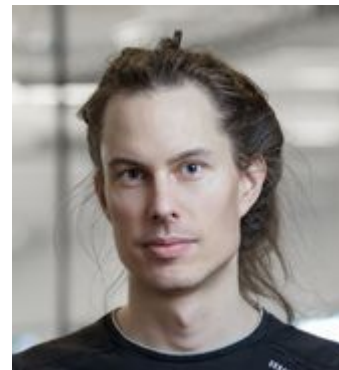
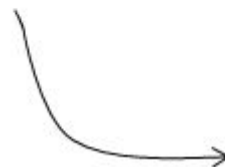


Подобрать такие вектора, чтобы чем чаще слова встречаются в одном контексте, тем они ближе.

word2vec

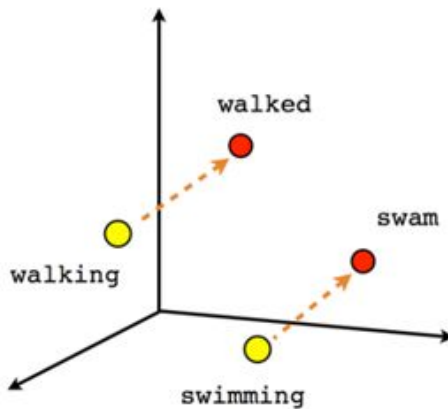
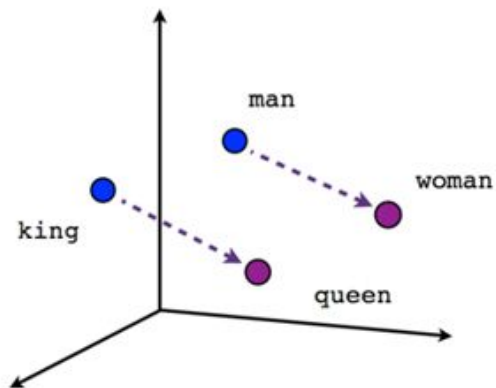


Tomas Mikolov
2013

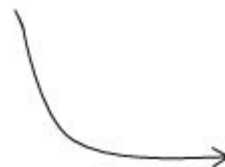


Подобрать такие вектора, чтобы чем чаще слова встречаются в одном контексте, тем они ближе.

word2vec

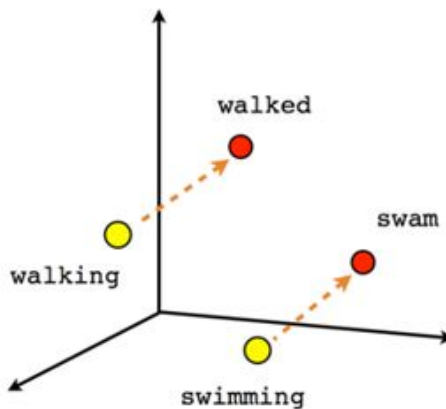
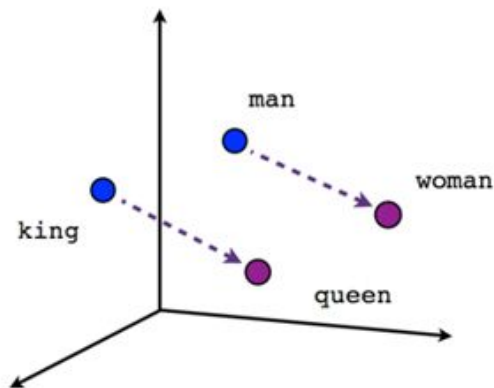


Tomas Mikolov
2013

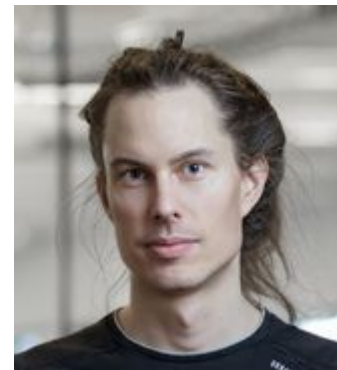
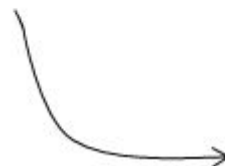


ЛОНДОН – АНГЛИЯ + РОССИЯ =

word2vec



Tomas Mikolov
2013



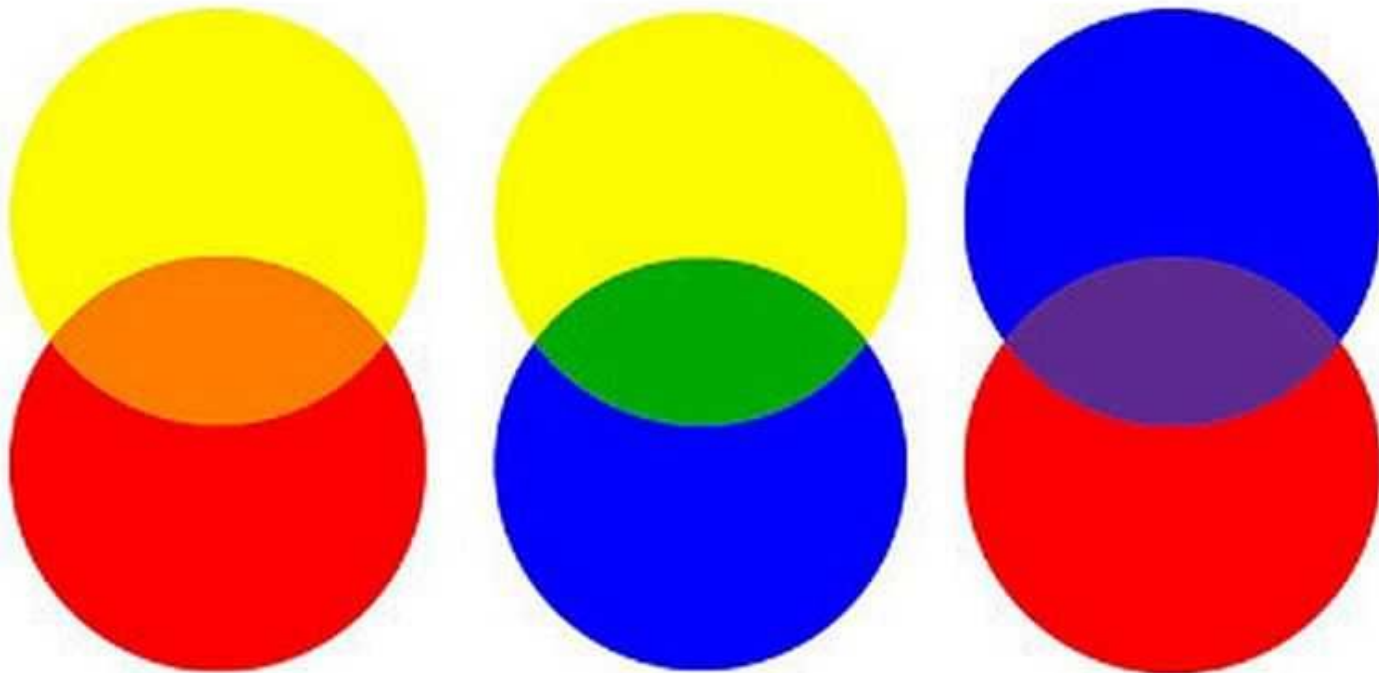
ЛОНДОН – АНГЛИЯ + РОССИЯ =

<https://rusvectors.org/ru/calculator/#>

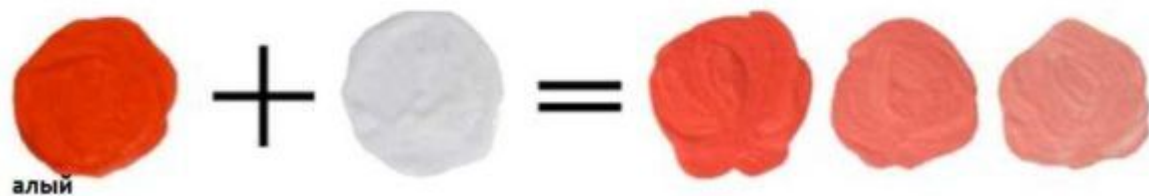
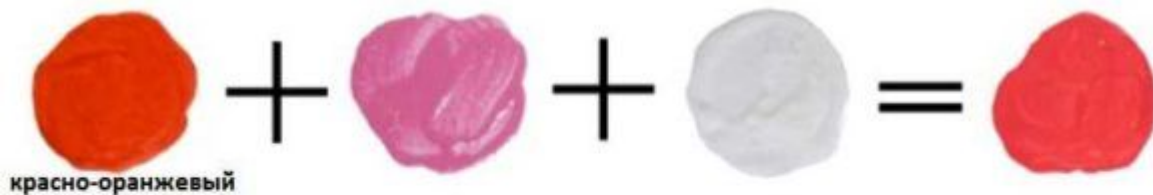
<https://rusvectors.org/en/#>

Как закодировать цвет числом?

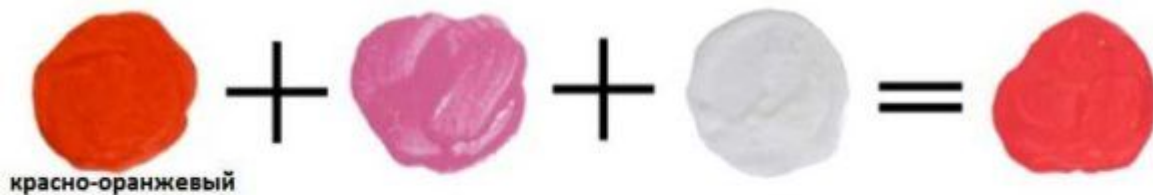
Все цвета можно получить смешиванием



Все цвета можно получить смешиванием



Все цвета можно получить смешиванием



Что получится?

А



Б



В



Что получится?

А



Б



В



Про цвета



- По две цифры на составляющую
- Две цифры образуют число в шестнадцатичной система счисления: помимо цифр 0, 1, 2, 3, ..., 9 есть a = 10, b = 11, ... , f = 15.

Про цвета



- 0f - это 15

Про цвета



- 0f - это 15
- 10_{16} - как записывается это число в привычной нам десятичной системе счисления

Про цвета



- 0f - это 15
- $10_{16} = 16_{10}$

Про цвета



- 0f - это 15
- $10_{16} = 16_{10}$
- $11_{16} = ?$

Про цвета



- 0f - это 15
- $10_{16} = 16_{10}$
- $11_{16} = 17_{10}$

Про цвета



- 0f - это 15
- $10_{16} = 16_{10}$
- $21_{16} = ?_{10}$

Про цвета



- 0f - это 15
- $10_{16} = 16_{10}$
- $21_{16} = 32_{10} + 1_{10}$

Про цвета



- Как выглядит самое большое двузначное число в шестнадцатеричной системе счисления?

Про цвета



- Как выглядит самое большое двузначное число в шестнадцатеричной системе счисления?

ff

Про цвета



$$ff_{16} = ?_{10}$$

Про цвета



$$\text{ff}_{16} = 255_{10}$$

Про цвета



Что больше?

33, 99 или ff?

Про цвета



Что больше?

33, 99 или **ff**?

Про цвета



Что больше?

af, 90 или f0?

Про цвета



Что больше?

af, 90 или f0?

Про цвета



Вернемся к цветам:

- 00 = отсутствие цвета
- ff = максимальная интенсивность

Про цвета



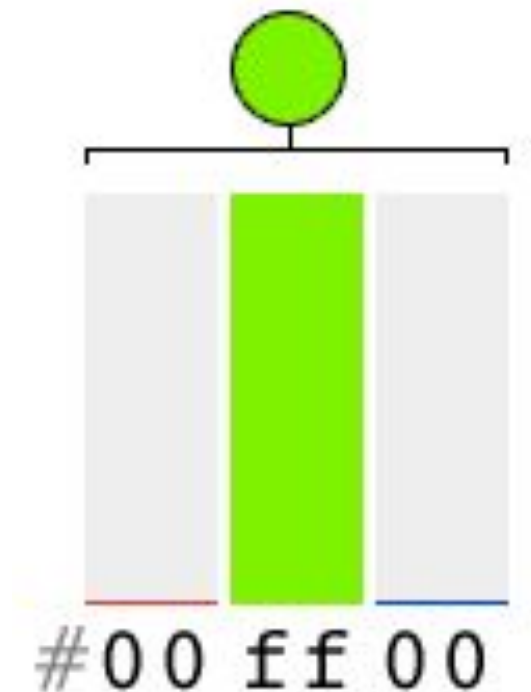
`#00ff00` - какой это цвет?

Про цвета



`#00ff00` - какой это цвет?

Зеленый



Про цвета



#ffff00 - какой это цвет?

Про цвета



#ffff00 - какой это цвет?

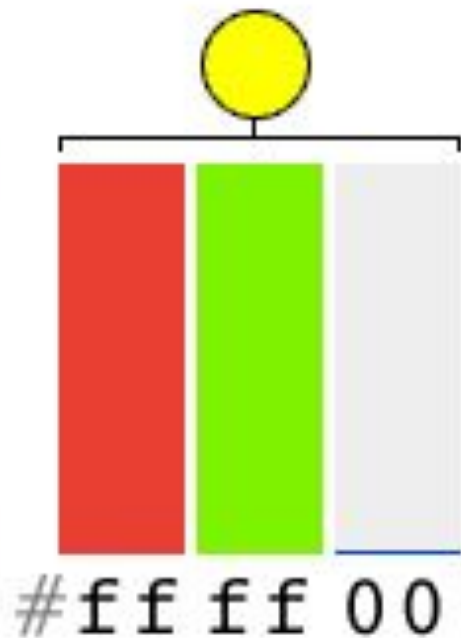
Желтый

Про цвета



`#ffff00` - какой это цвет?

Желтый



Про цвета

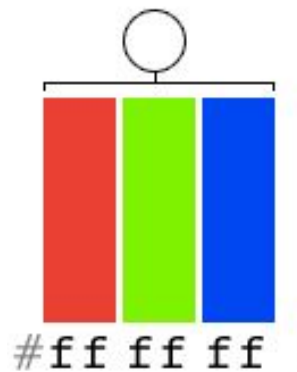
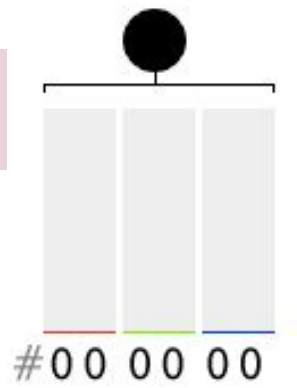


Как кодируется белый?

Про цвета



Как кодируется белый?



Животные



#3caa3c

#834d18

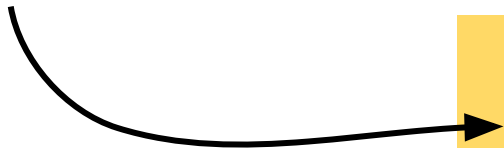
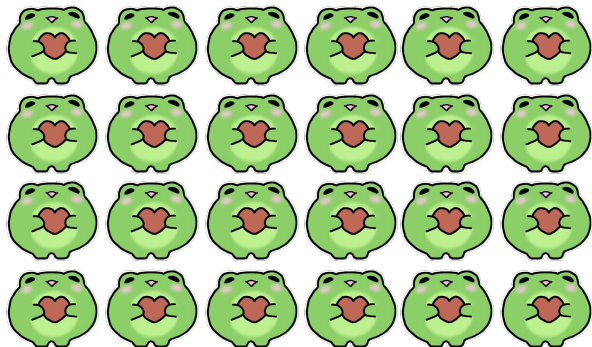
#008080



Животные



Влюбленные жабы



#3caa3c

#834d18

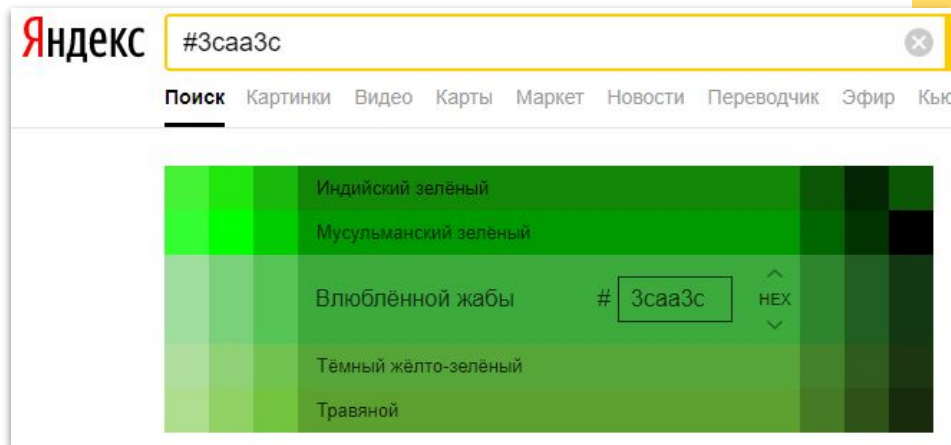
#008080



Животные



Влюбленные жабы



#3caa3c

#834d18

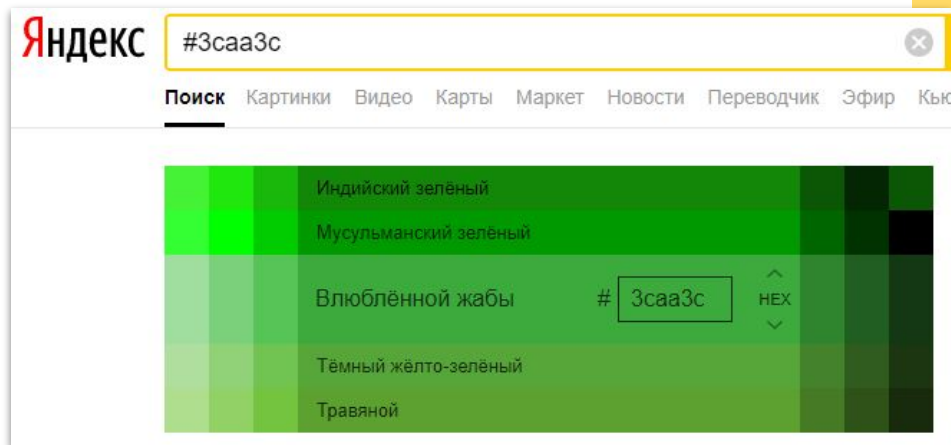
#008080



Животные



Влюбленные жабы



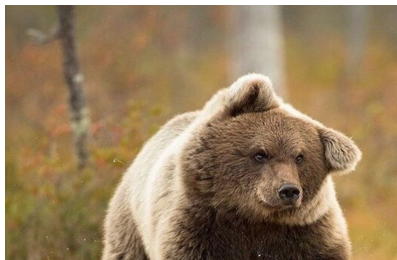
#3caa3c

#834d18

#008080



Животные



Цвет медвежьего
ушка

#834D18
(131,77,24)

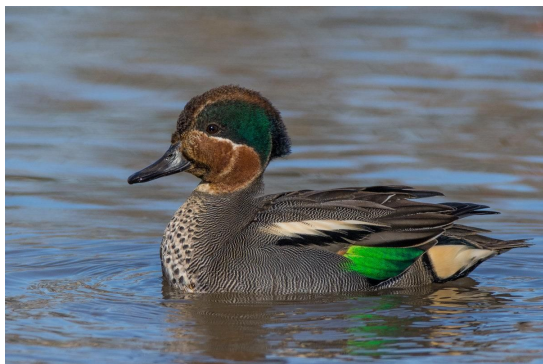
#3caa3c

#834d18

#008080



Животные



Утка "Чирок"



#008080

#3caa3c
#834d18
#008080



<http://www.hexinvaders.com/>

- Вы будете спасать планету.
- Тройки разноцветных пришельцев атакуют нас.
- У вас есть мощный цветовой заряд. Если им попасть по соответствующему пришельцу, то вся тройка будет побеждена.



Выберите
пришельца
данного
цвета

Задание №1

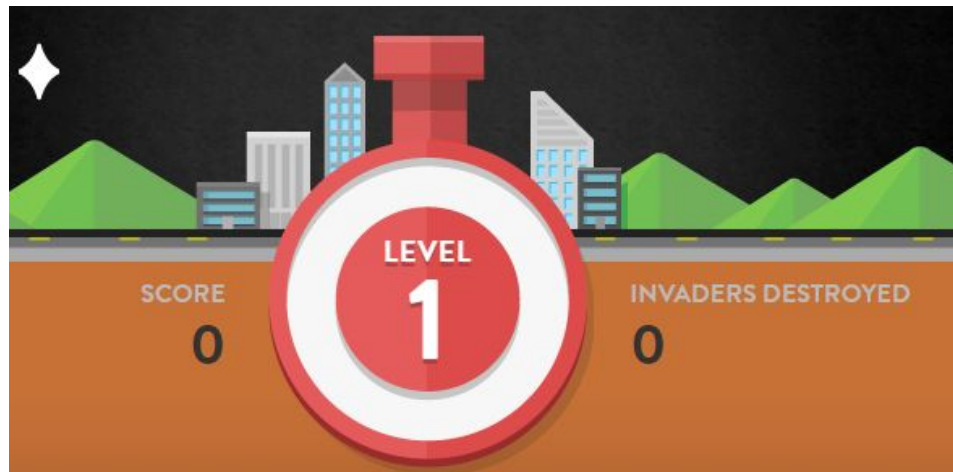
- Перейдите на
- Играйте, пока не проиграете или вас не остановит преподаватель
- Напишите ниже ваш лучший результат (не для сравнения с кем-то, просто для истории)

Результат:

LEVEL =

Score =

Invaders destroyed =

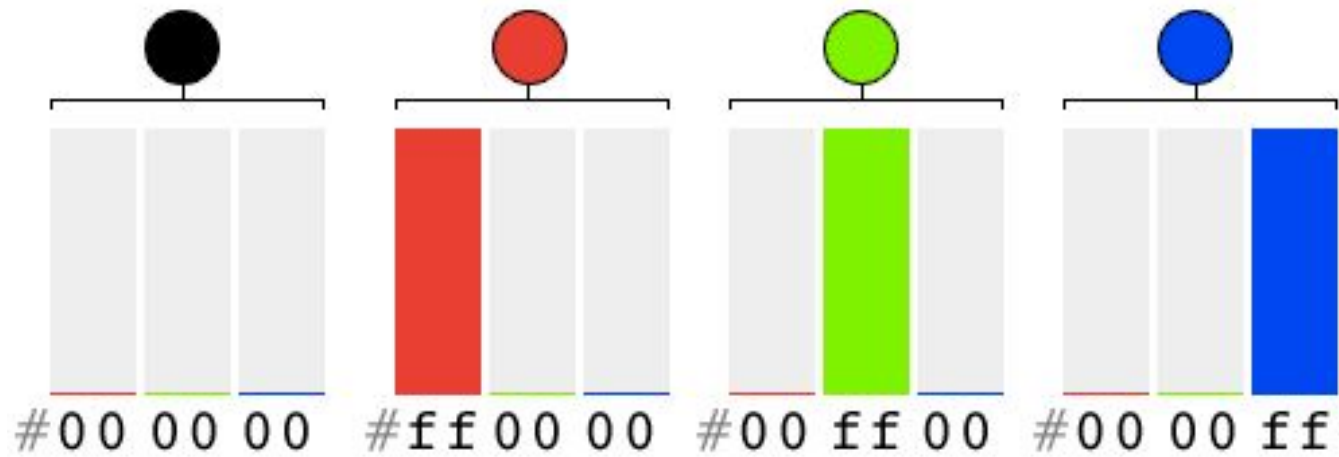


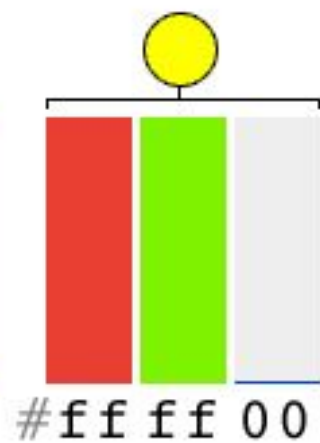
Опционально

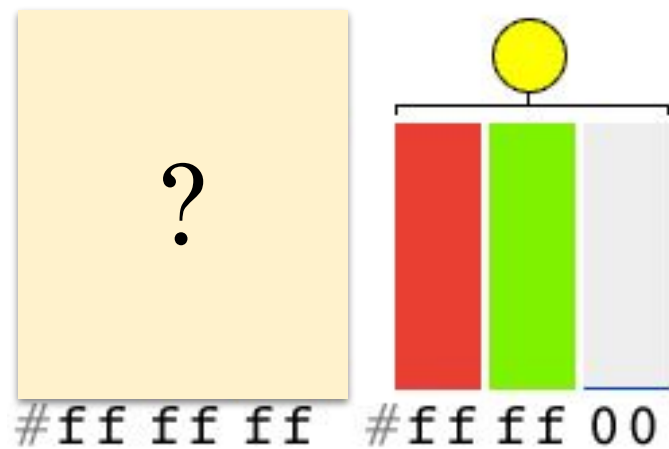
Про цвета

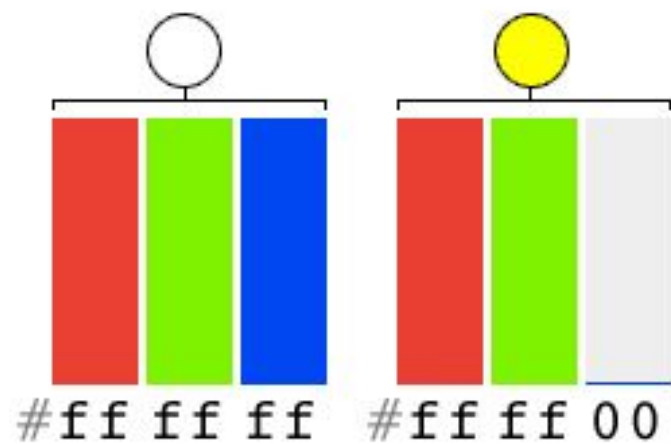
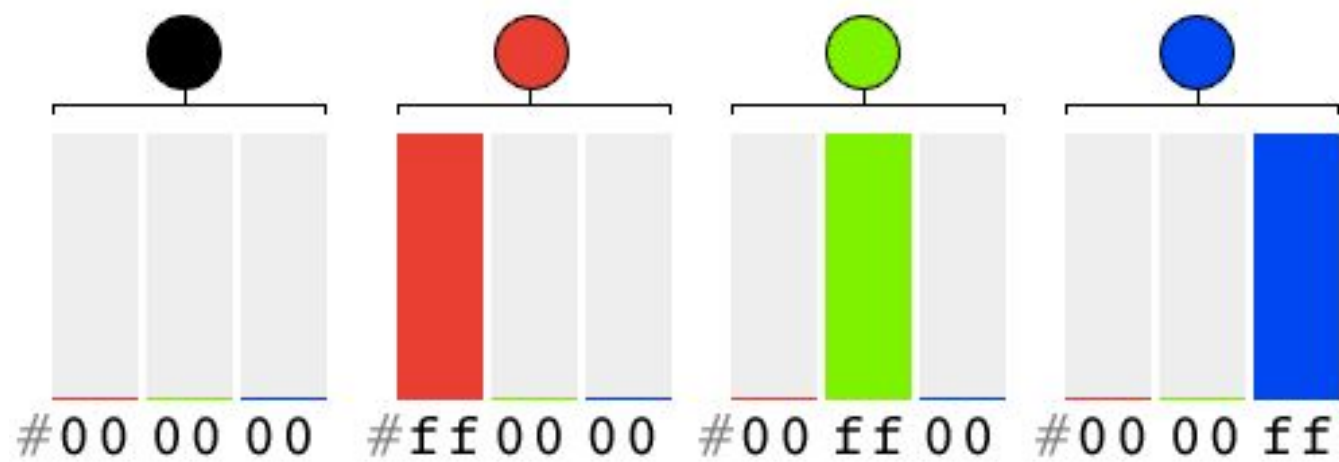


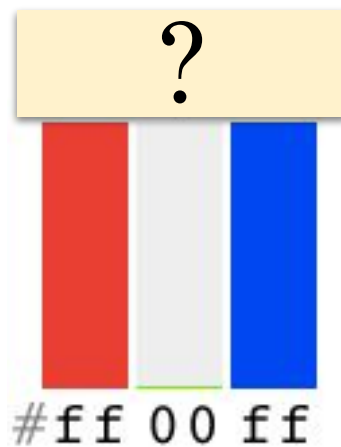
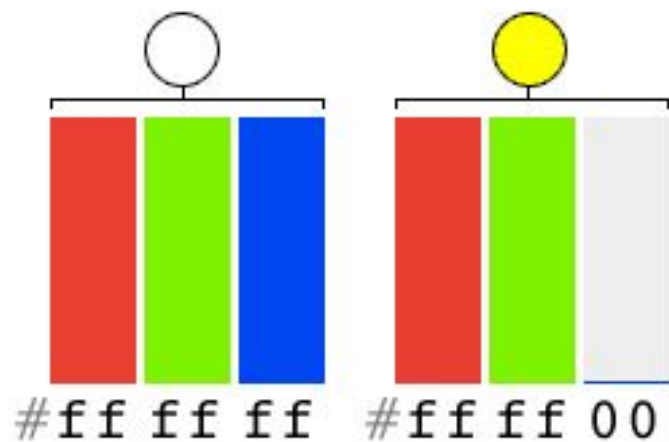
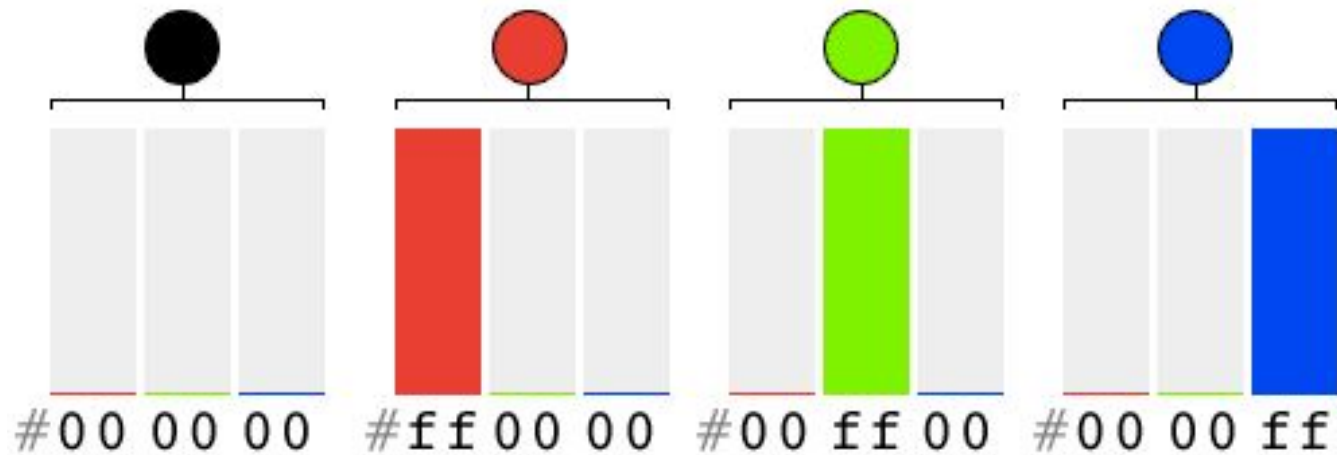
Как считаете, можно считать черный и белый цветами?
Почему их нет в радуге?

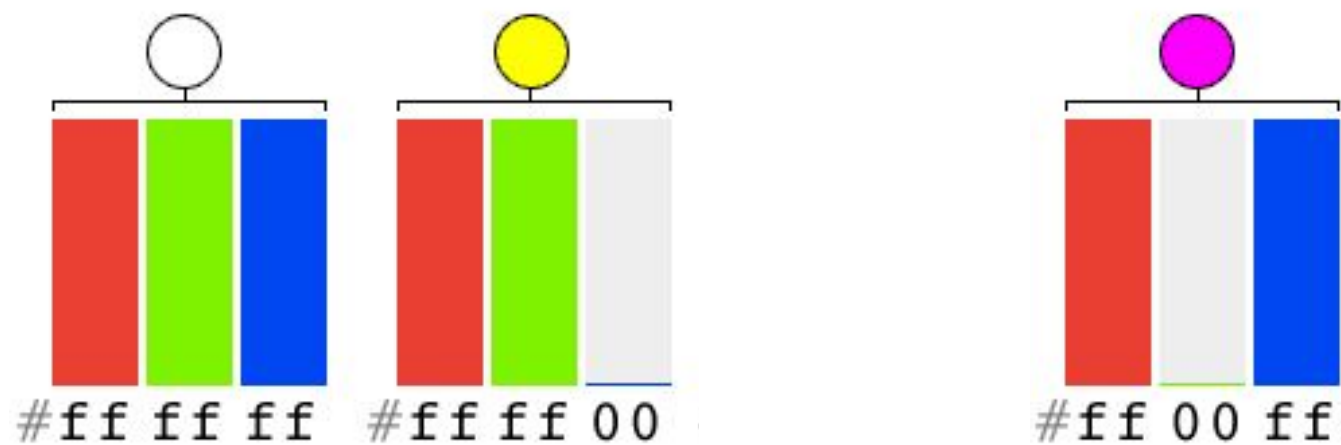


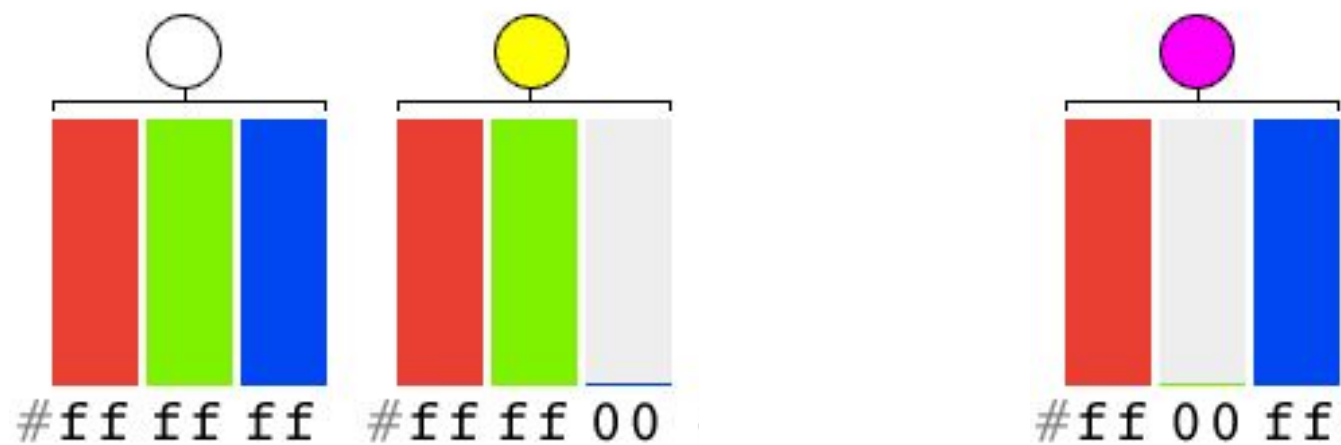


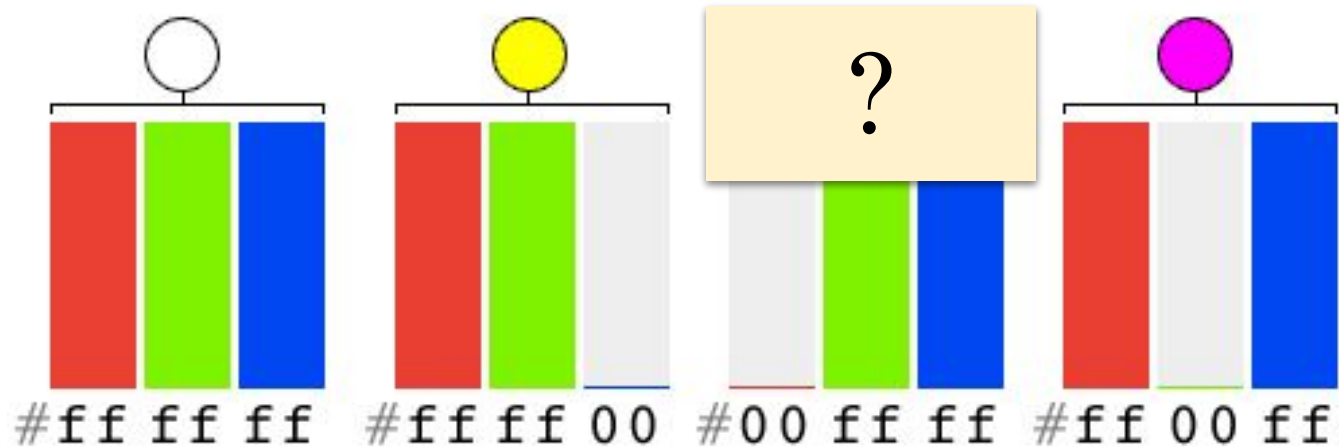
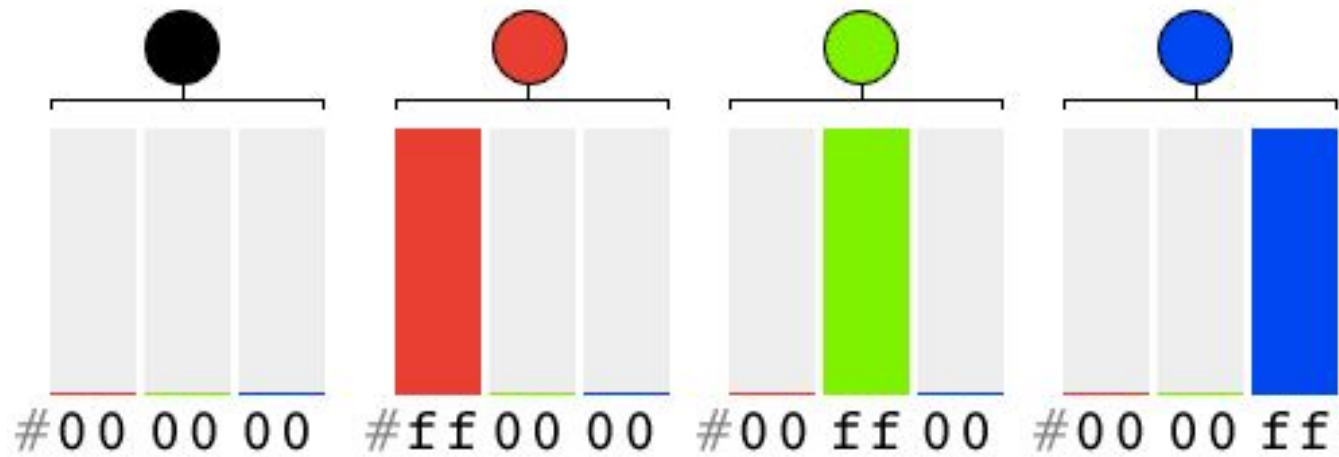


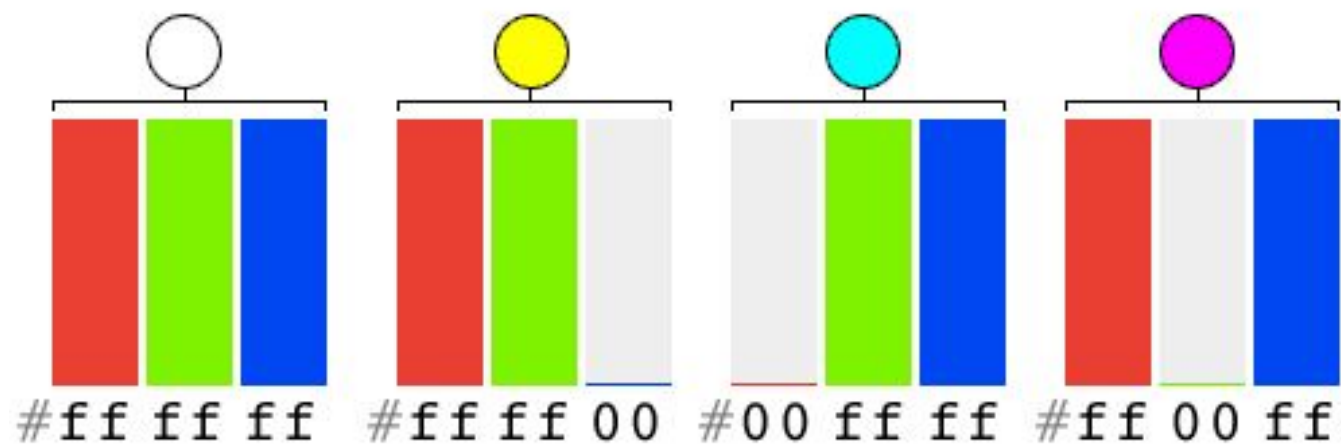
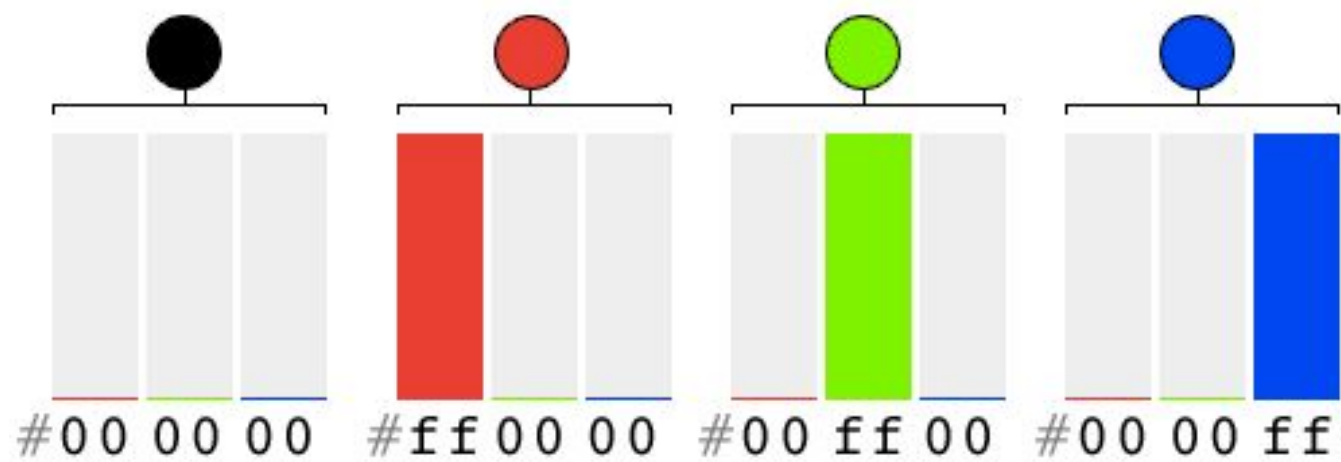












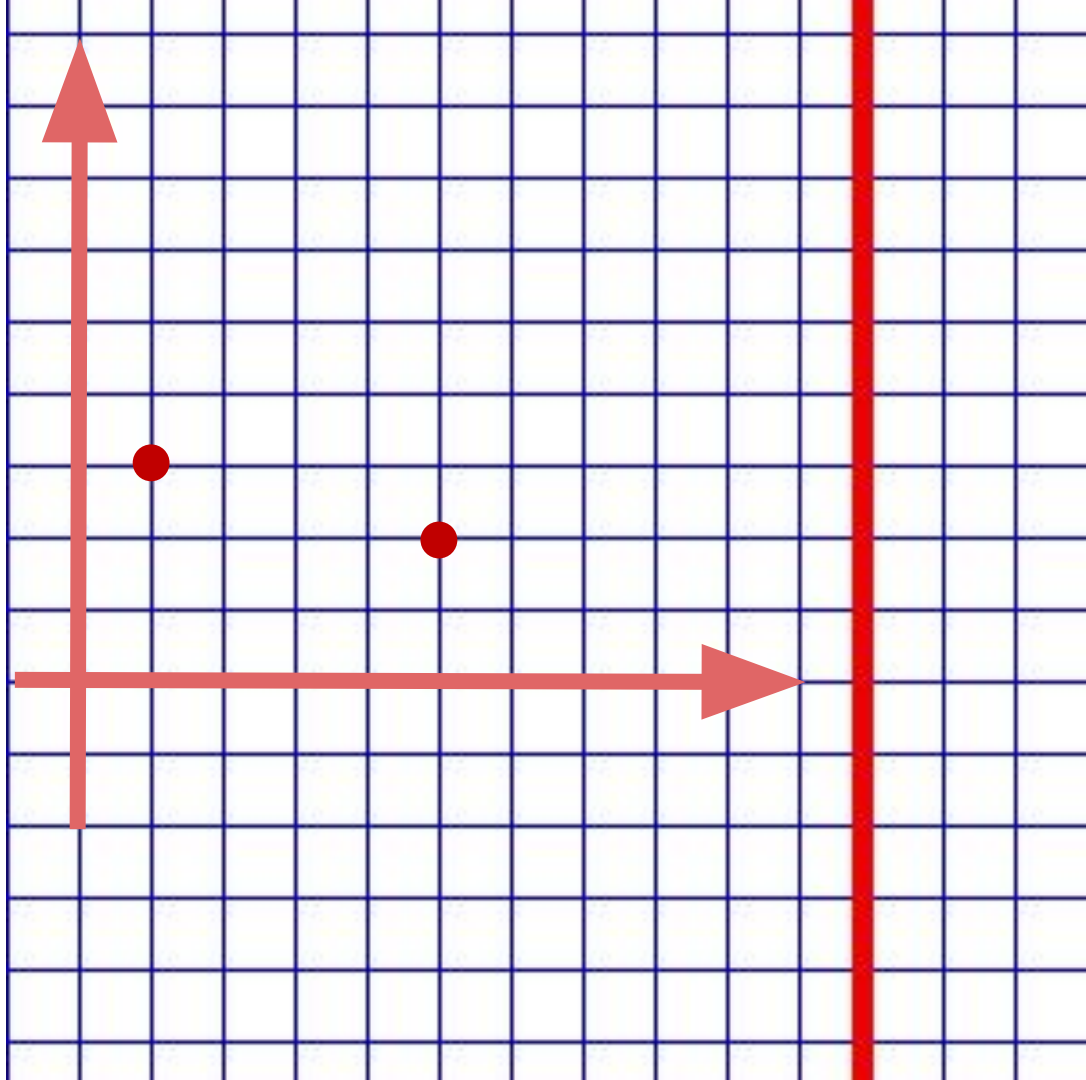
Как оперировать?

Цель

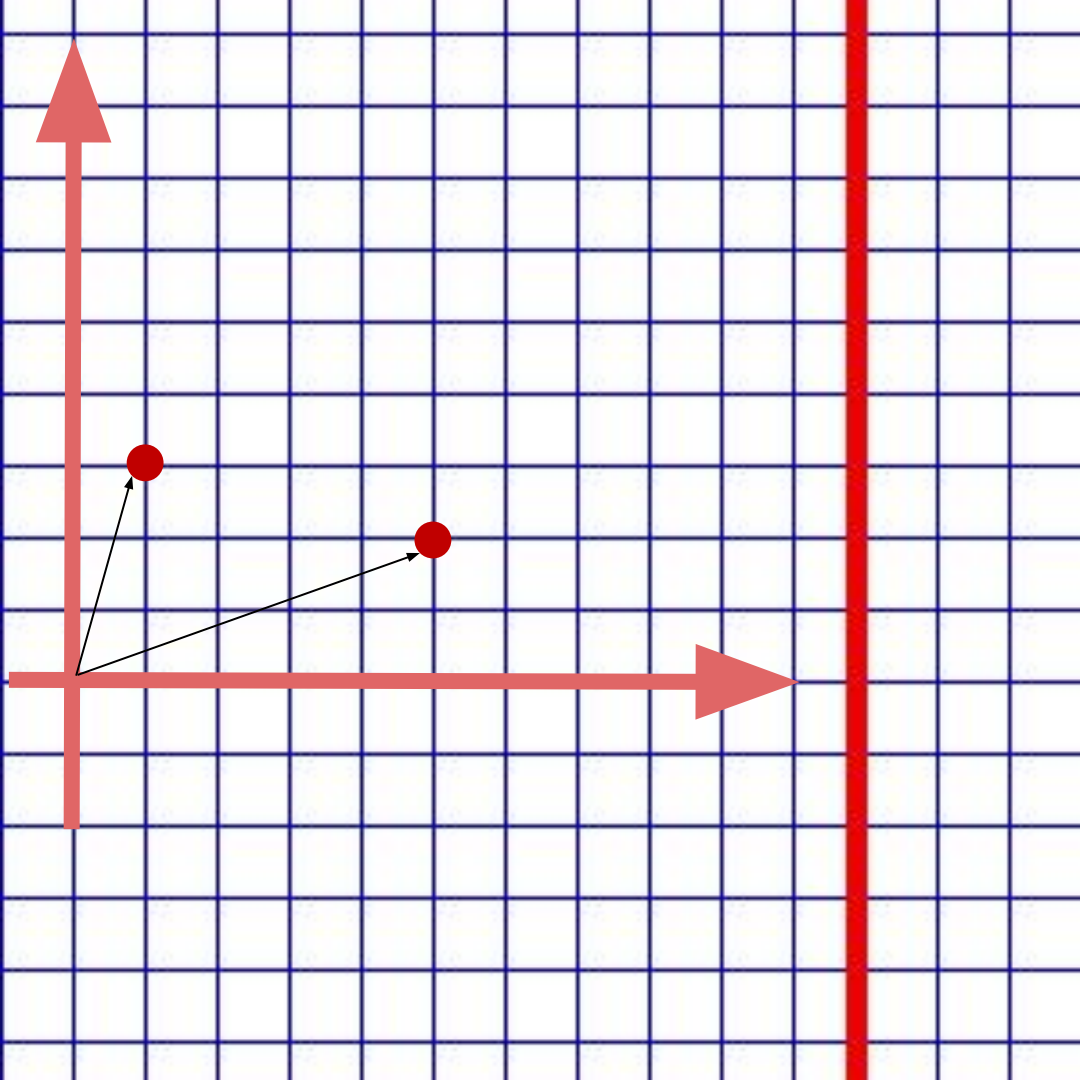


- Научиться кодировать разные объекты “хорошо”
 - Числом
 - Есть возможность сравнения разных объектов
 - Разница между объектами интерпретируема
 - “Скрещивать объекты”

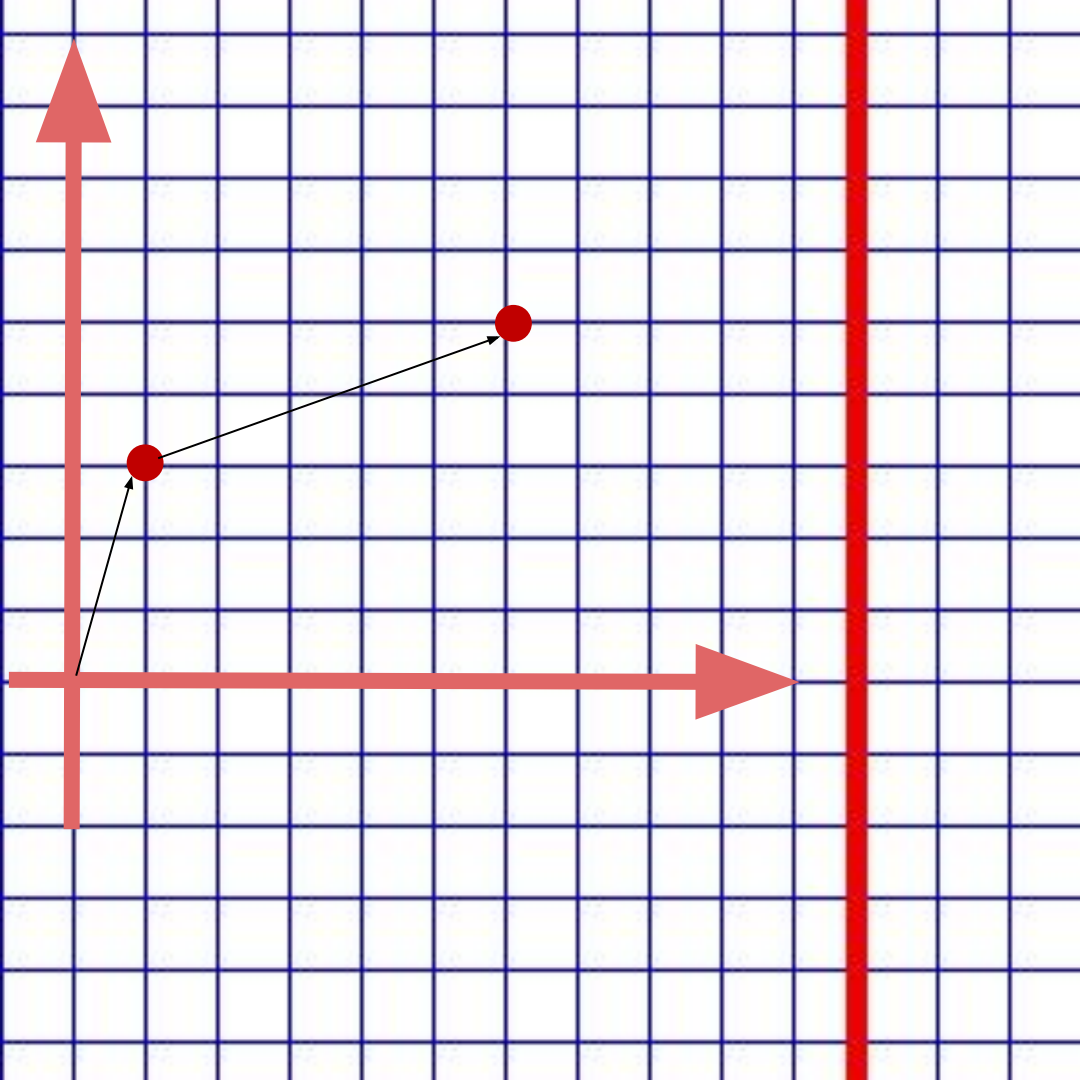
С точкой непонятно, что
делать



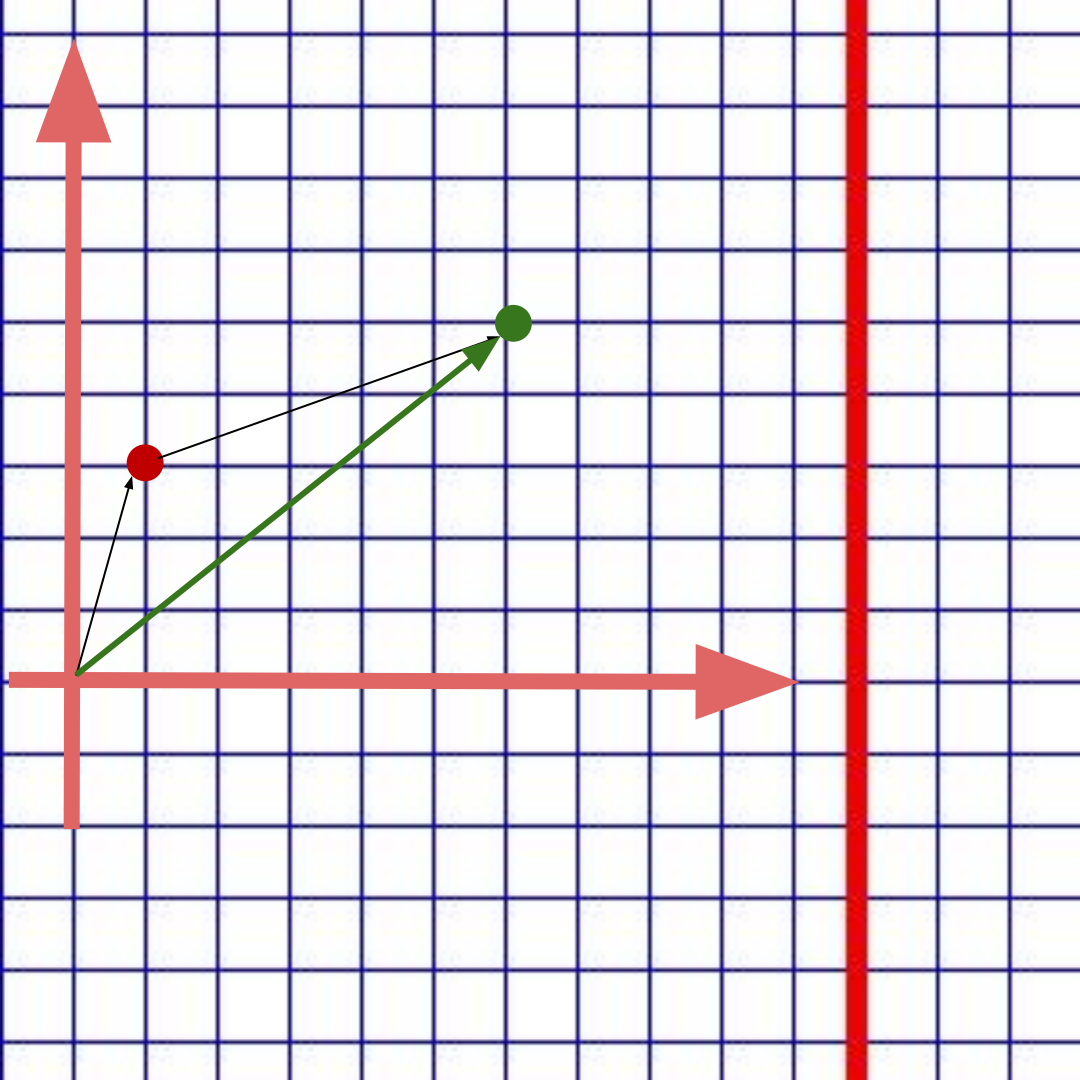
С точкой непонятно, что
делать



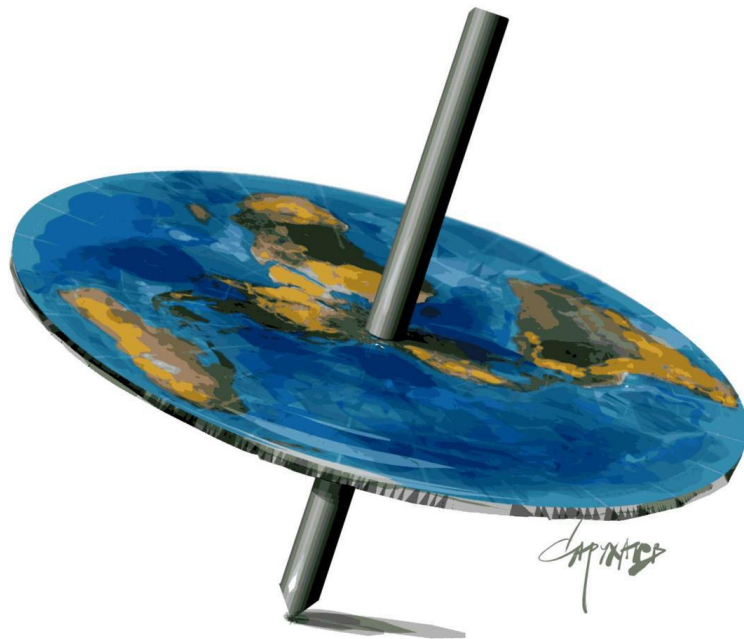
С точкой непонятно, что
делать



С точкой непонятно, что
делать



В ближайшее время речь будет идти
только о *двумерном пространстве*



Физика

?

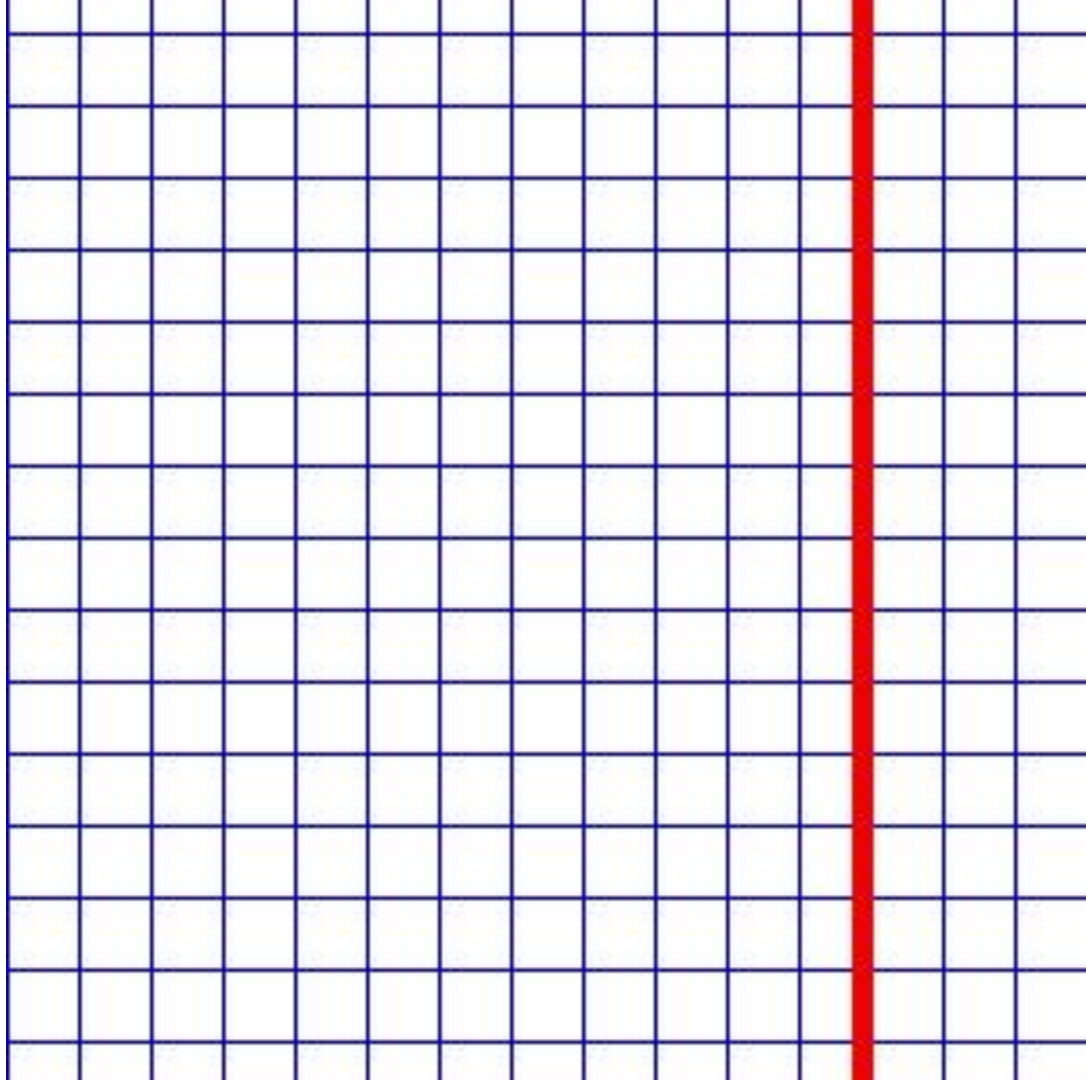
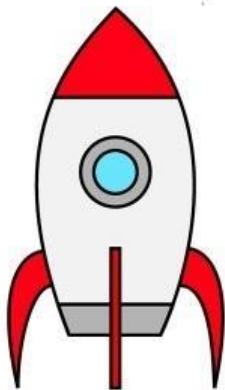
DS

?

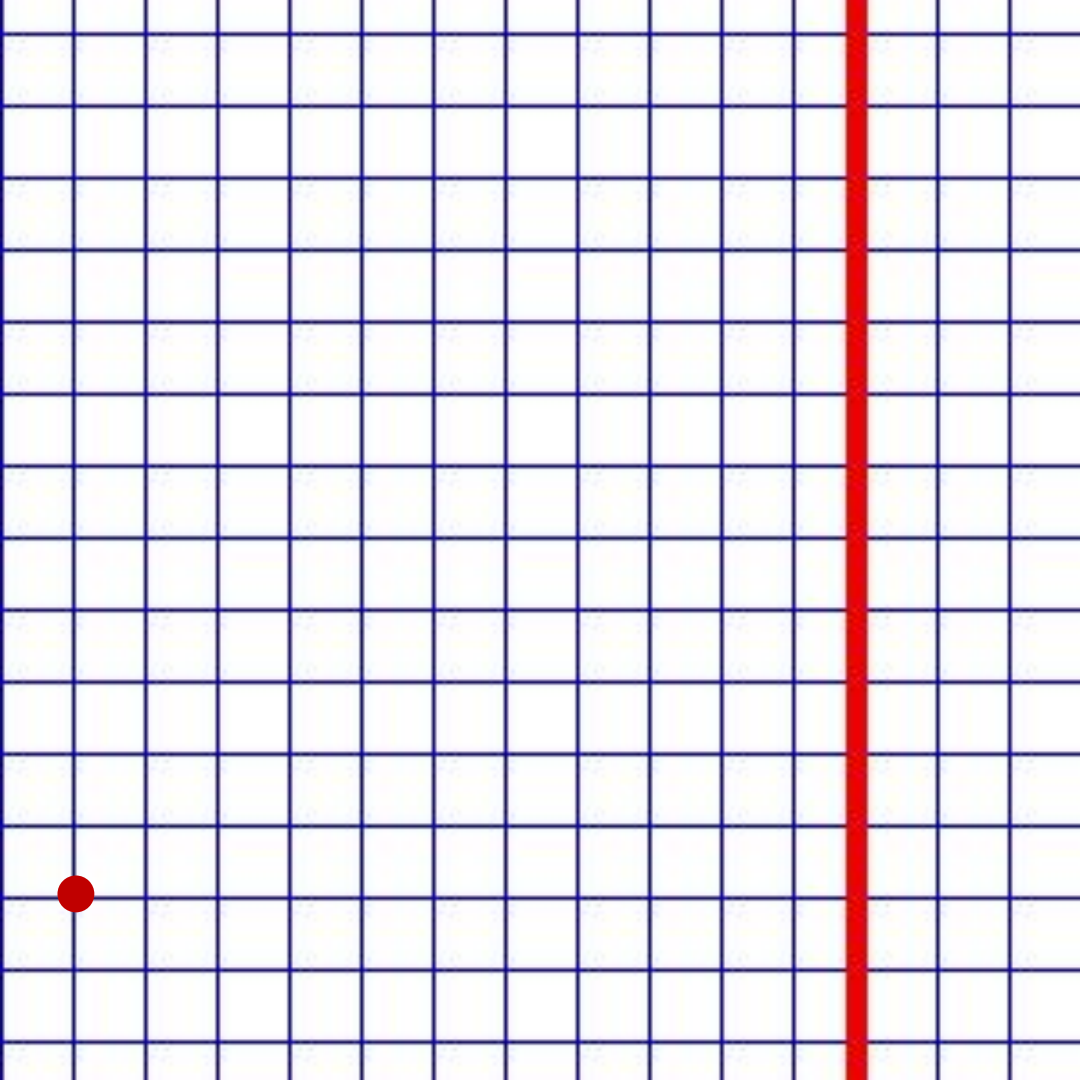
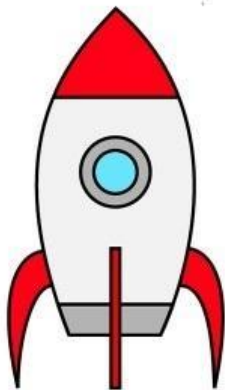
Math

?

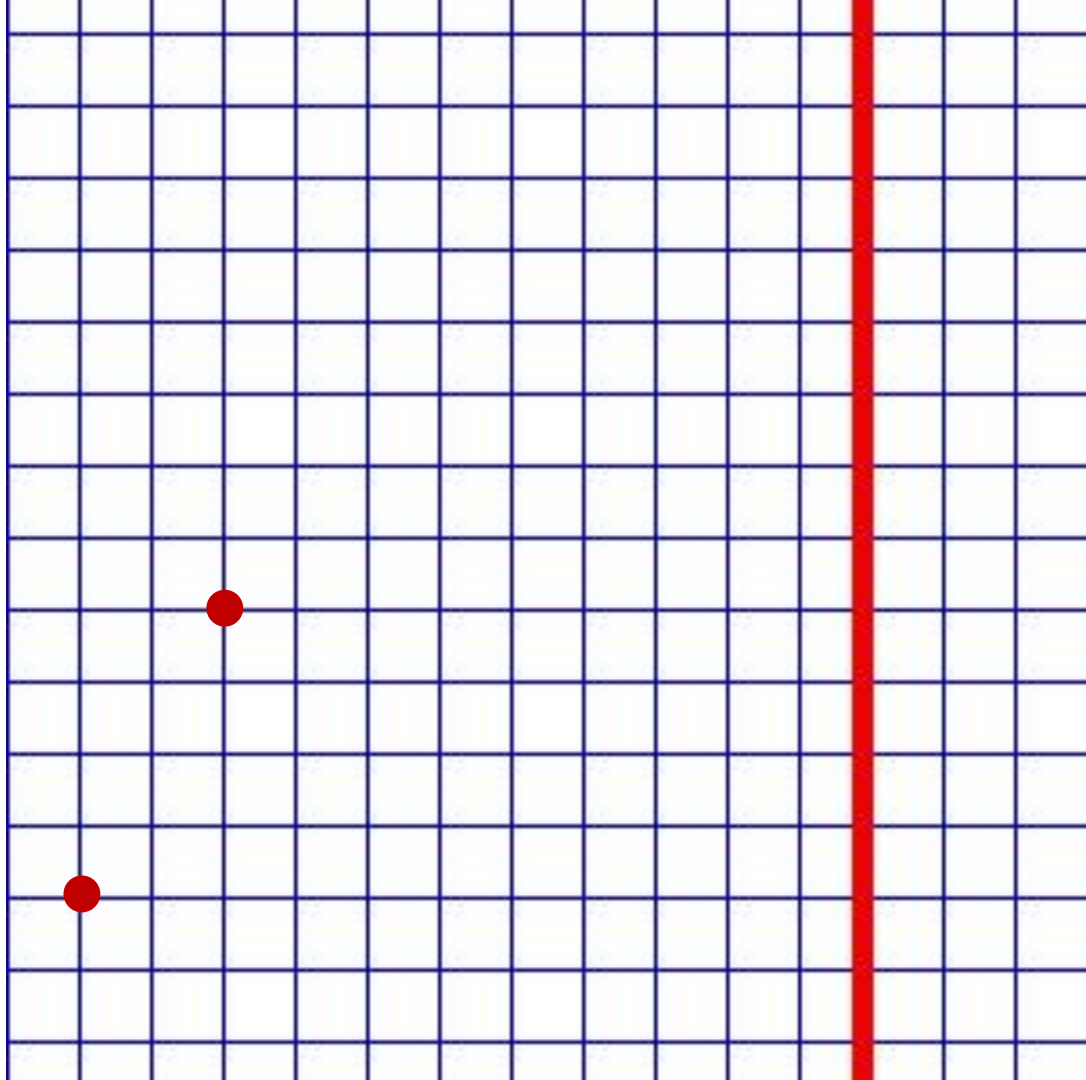
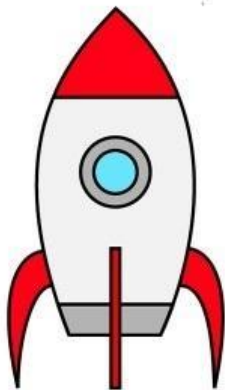
Вектор – это способ задать направление.



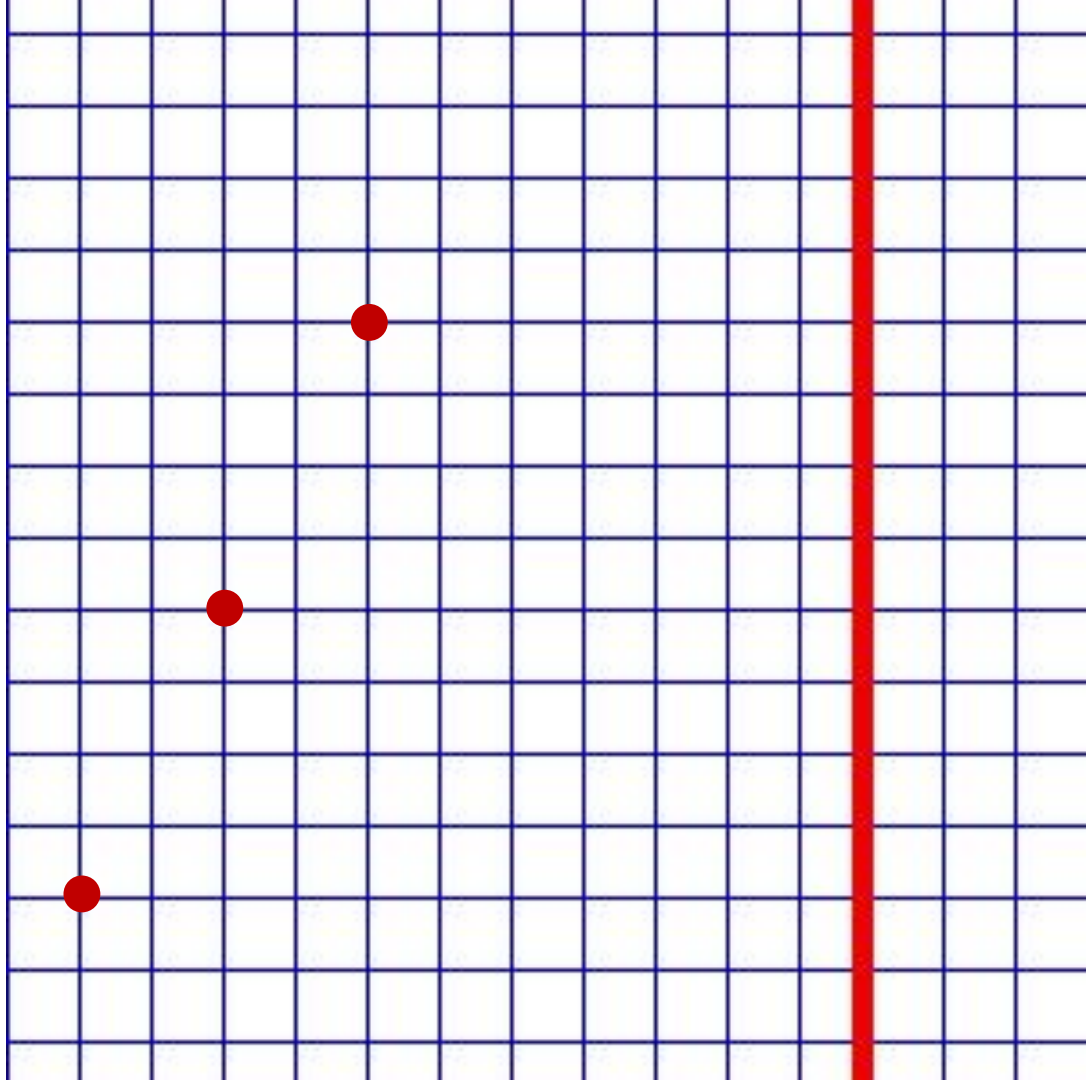
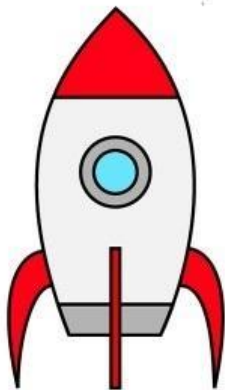
Вектор – это способ задать направление.



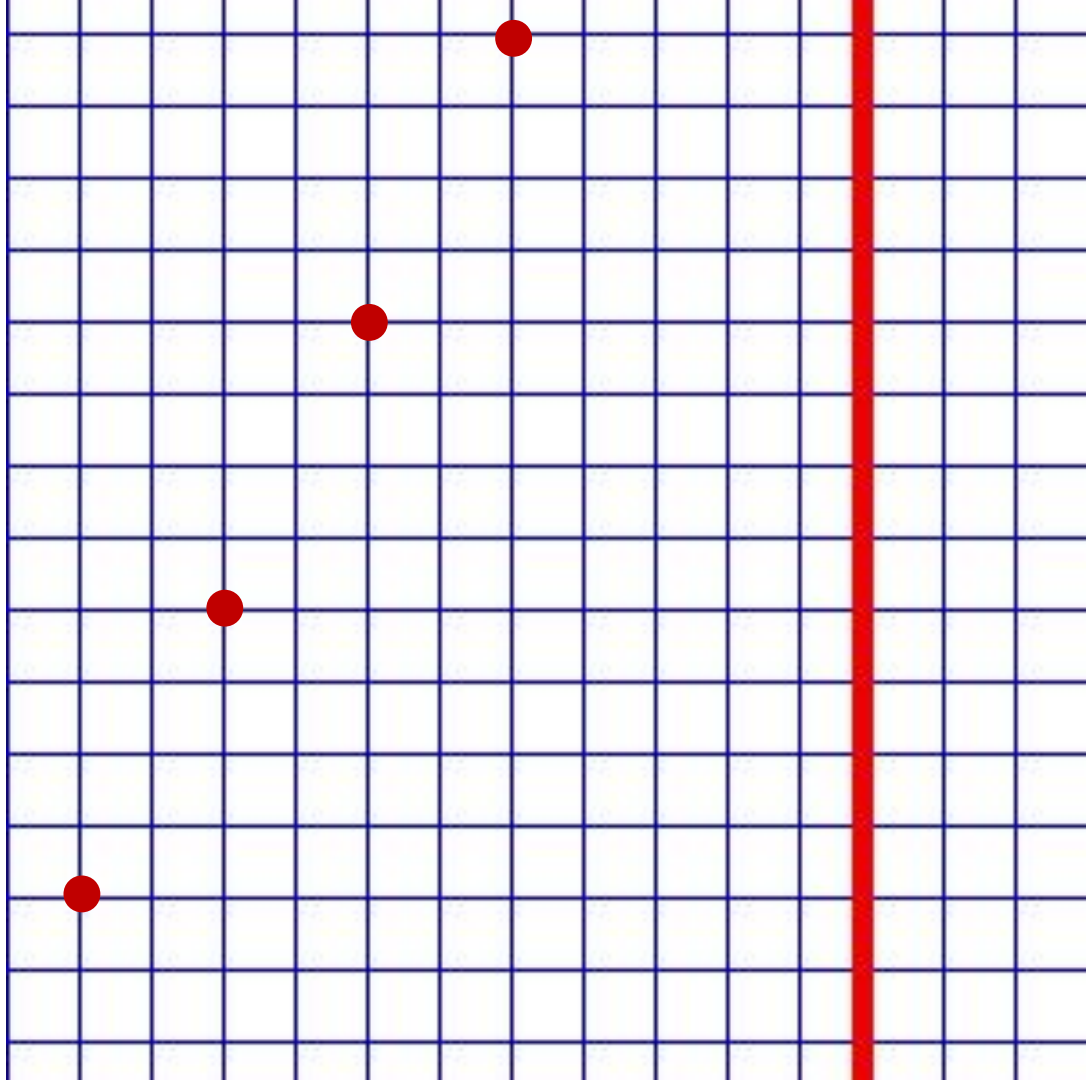
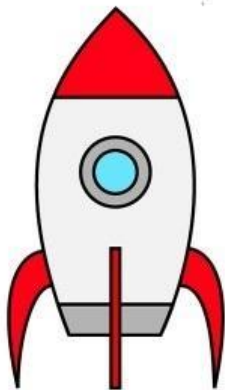
Вектор – это способ задать направление.



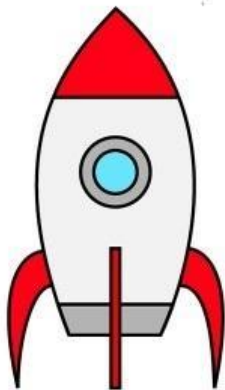
Вектор – это способ задать направление.



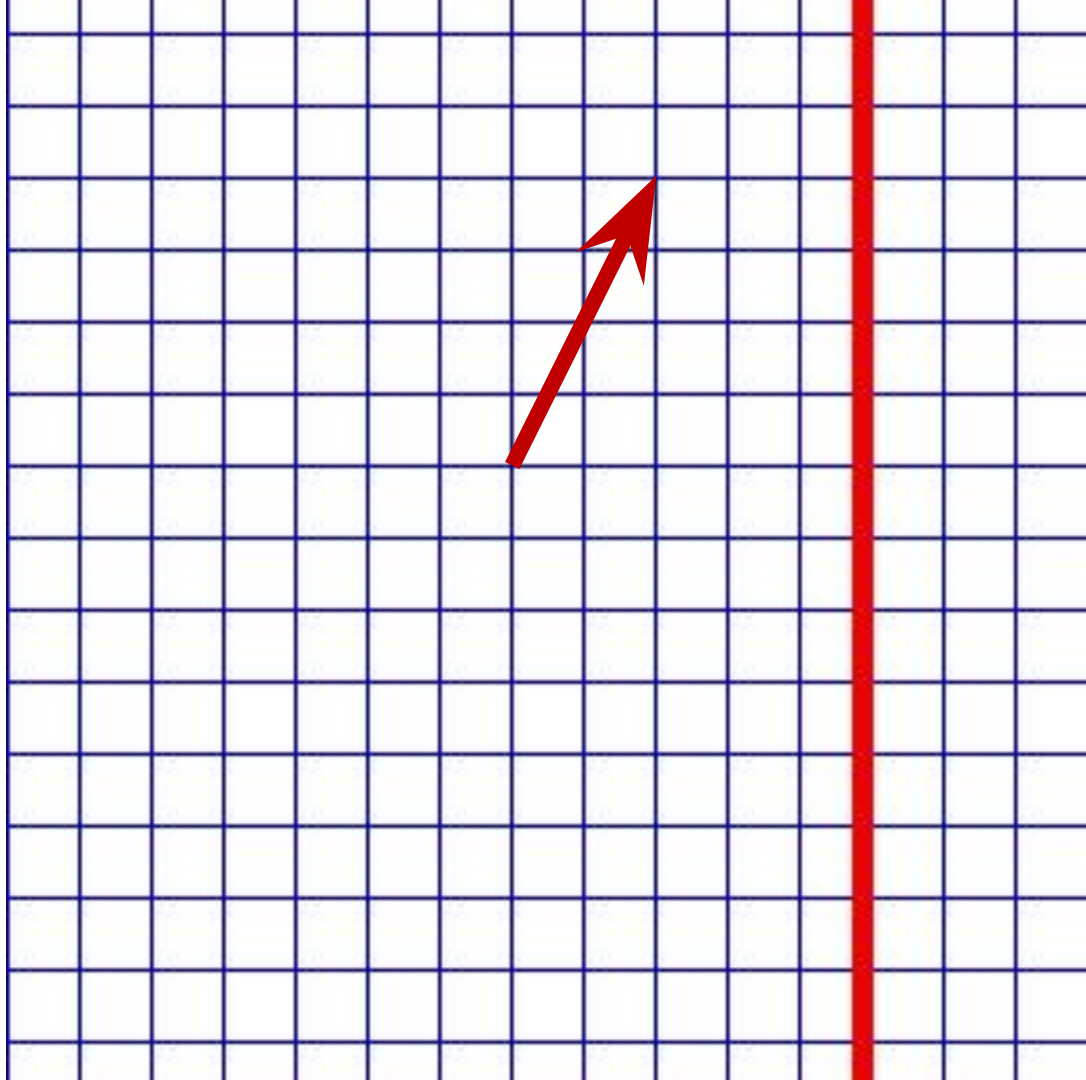
Вектор – это способ задать направление.



Вектор – это способ задать направление.



Вектор «не прикреплён» к
конкретной точке
пространства.
Он задает только
направление.



Способы описания вектора

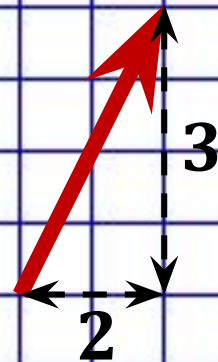
*... когда нет возможности
передать изображение*



Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали:

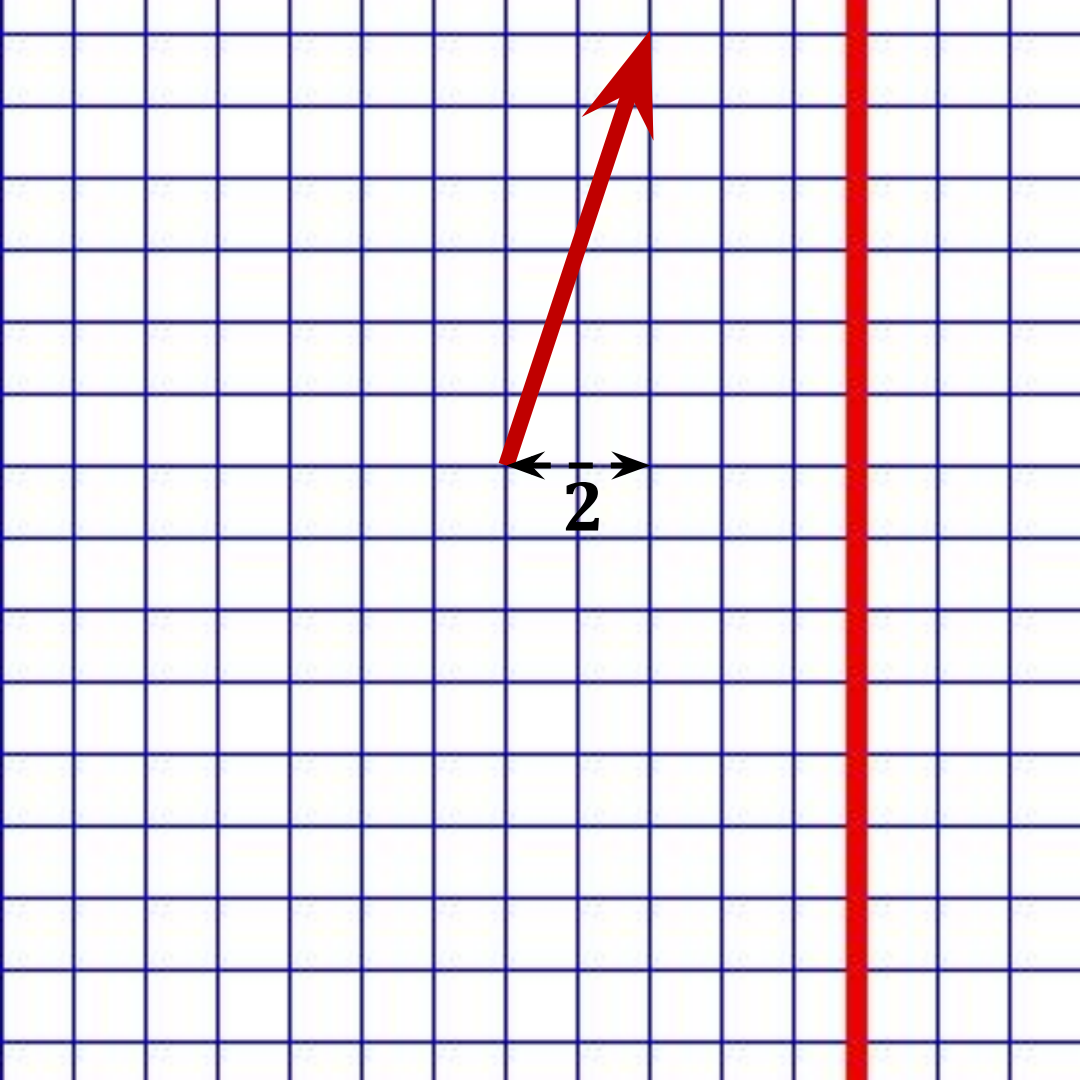
$(2, 3)$



Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали:

Заметим, что одного числа для описания мало



Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали:

Заметим, что одного числа для описания мало



Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали:

А у такого вектора?



Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали:

$(-4, 0)$



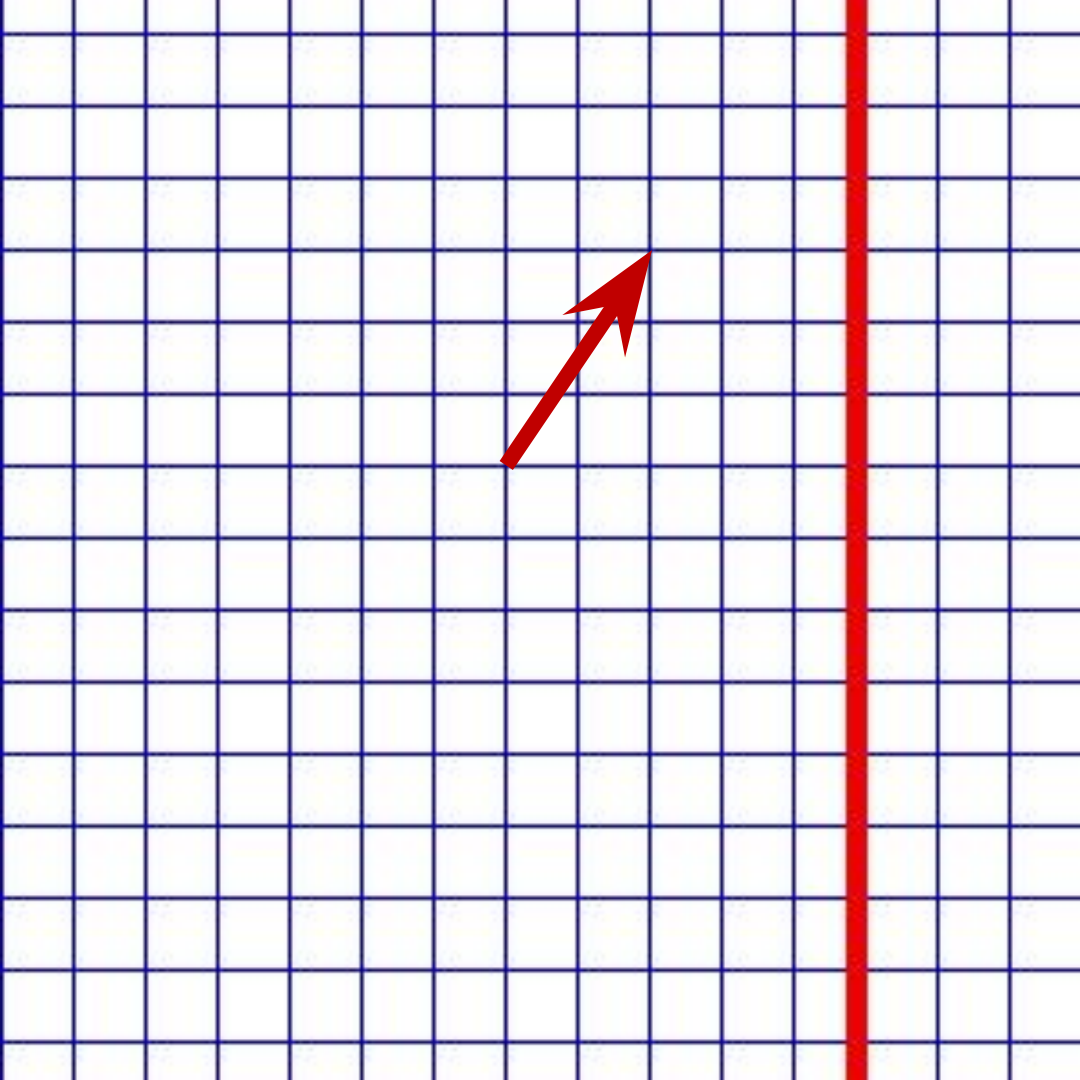
Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали:

$(2, 3)$

Чтобы отличать от точки:

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$



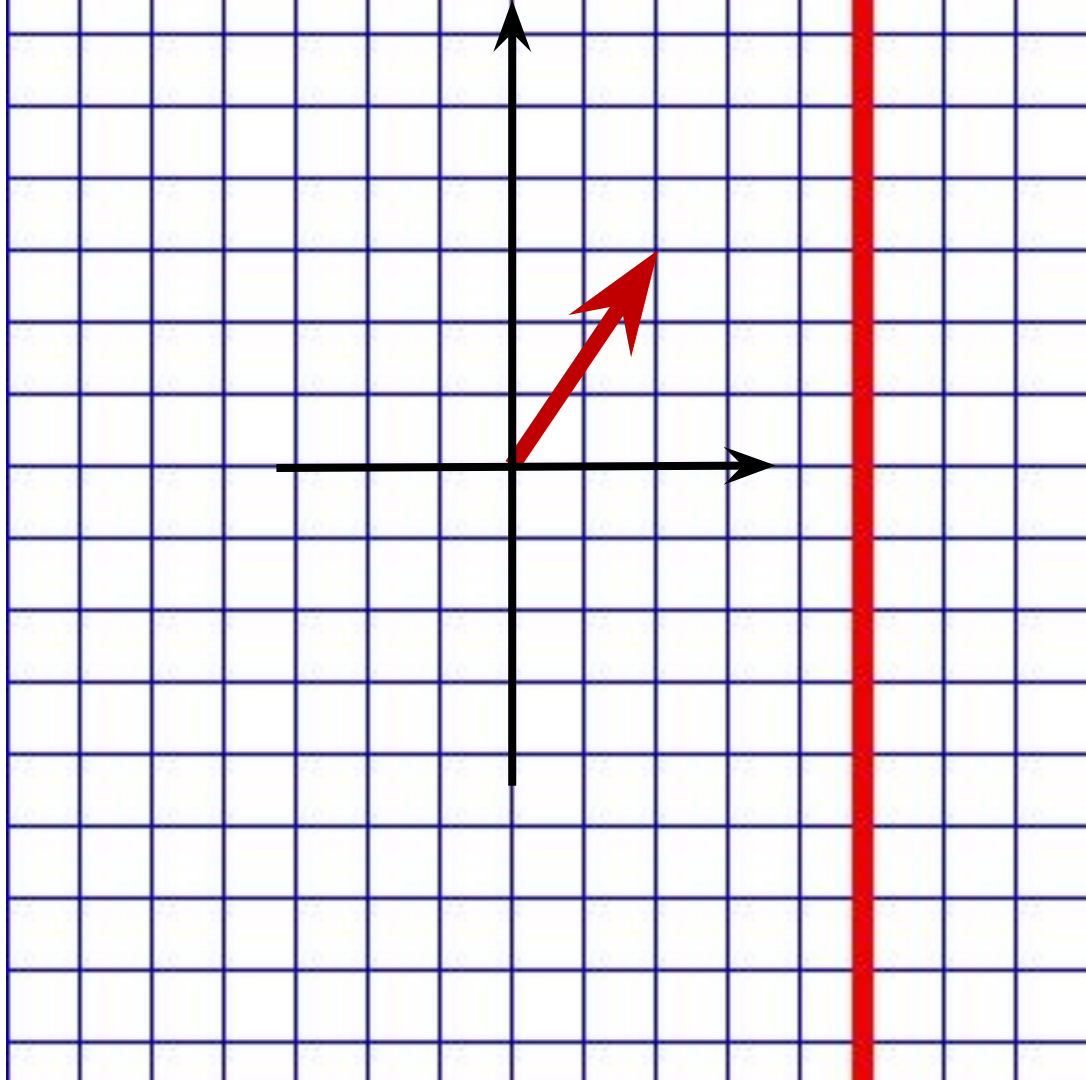
Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали:

$(2, 3)$

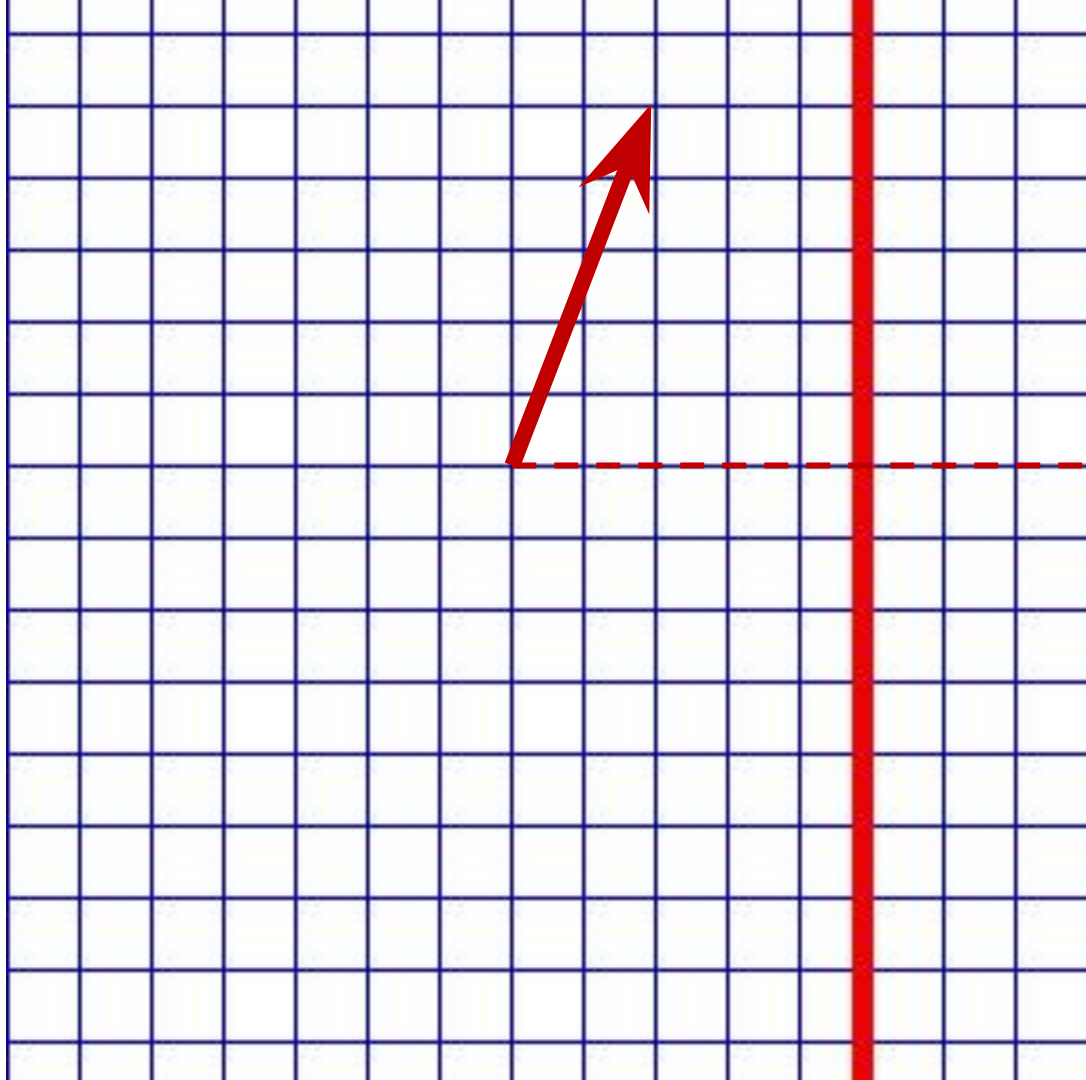
Чтобы отличать от точки:

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$



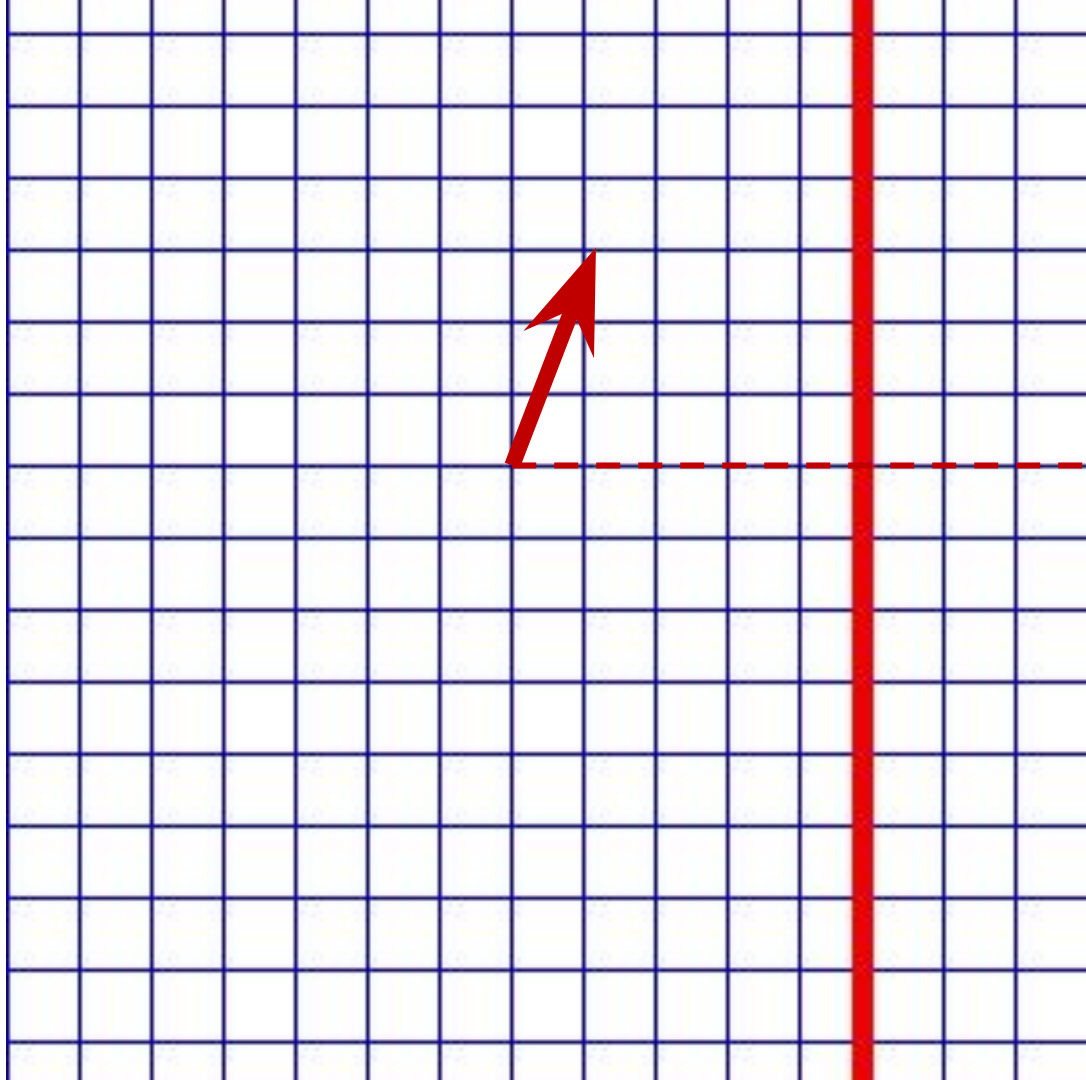
Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали.
2. Угол ...



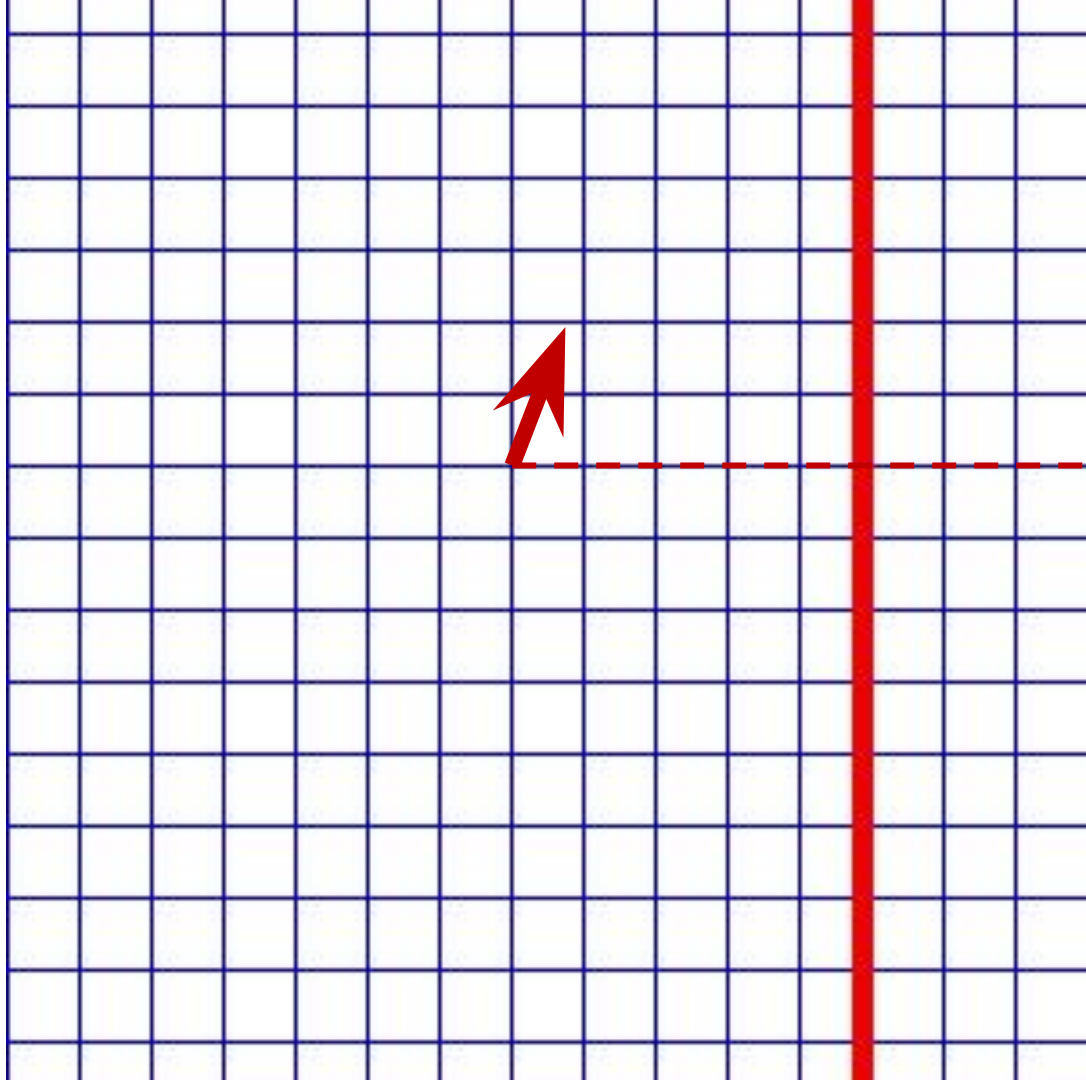
Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали.
2. Угол ...



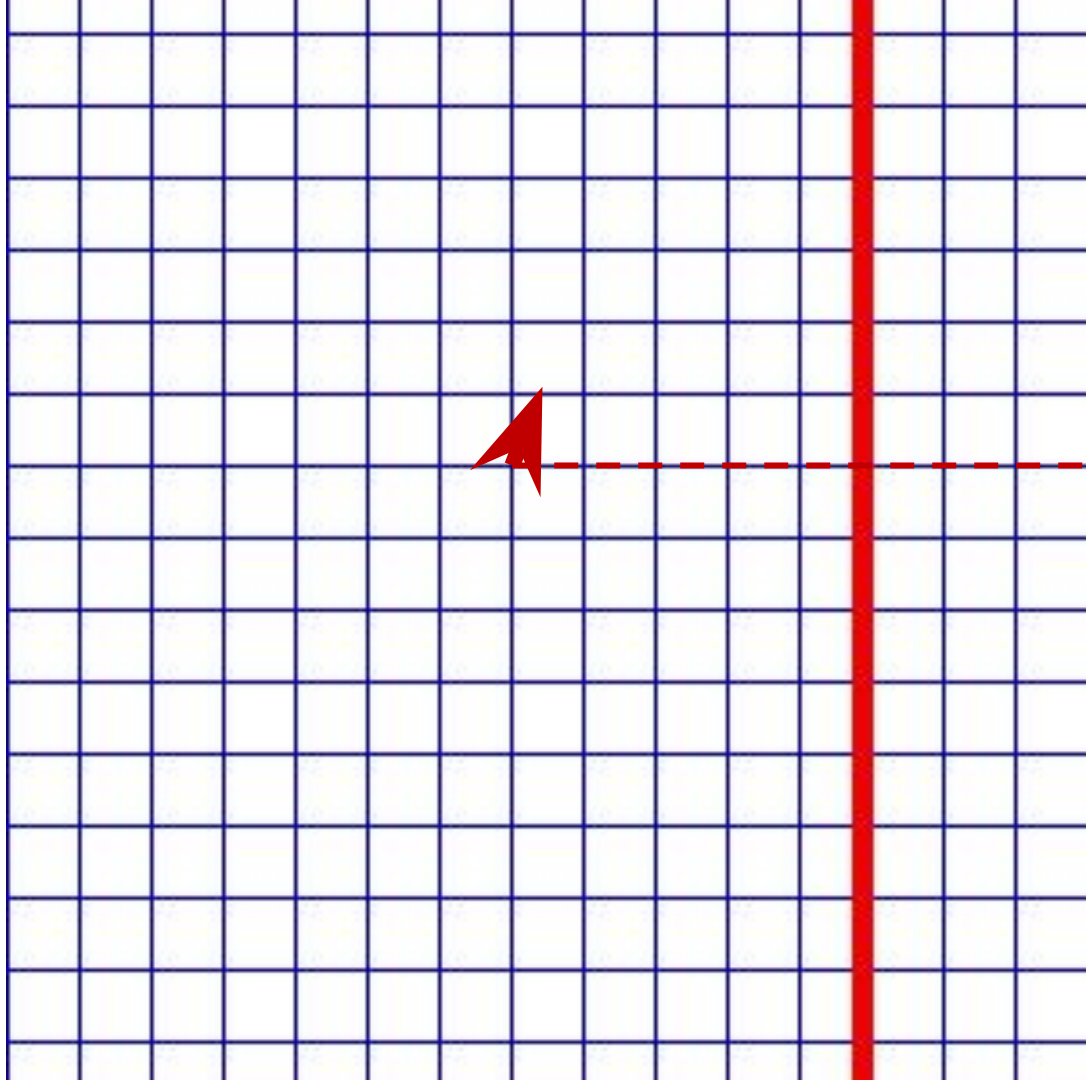
Способы описания вектора

1. На сколько сдвигается по горизонтали и по вертикали.
2. Угол ...



Способы описания вектора

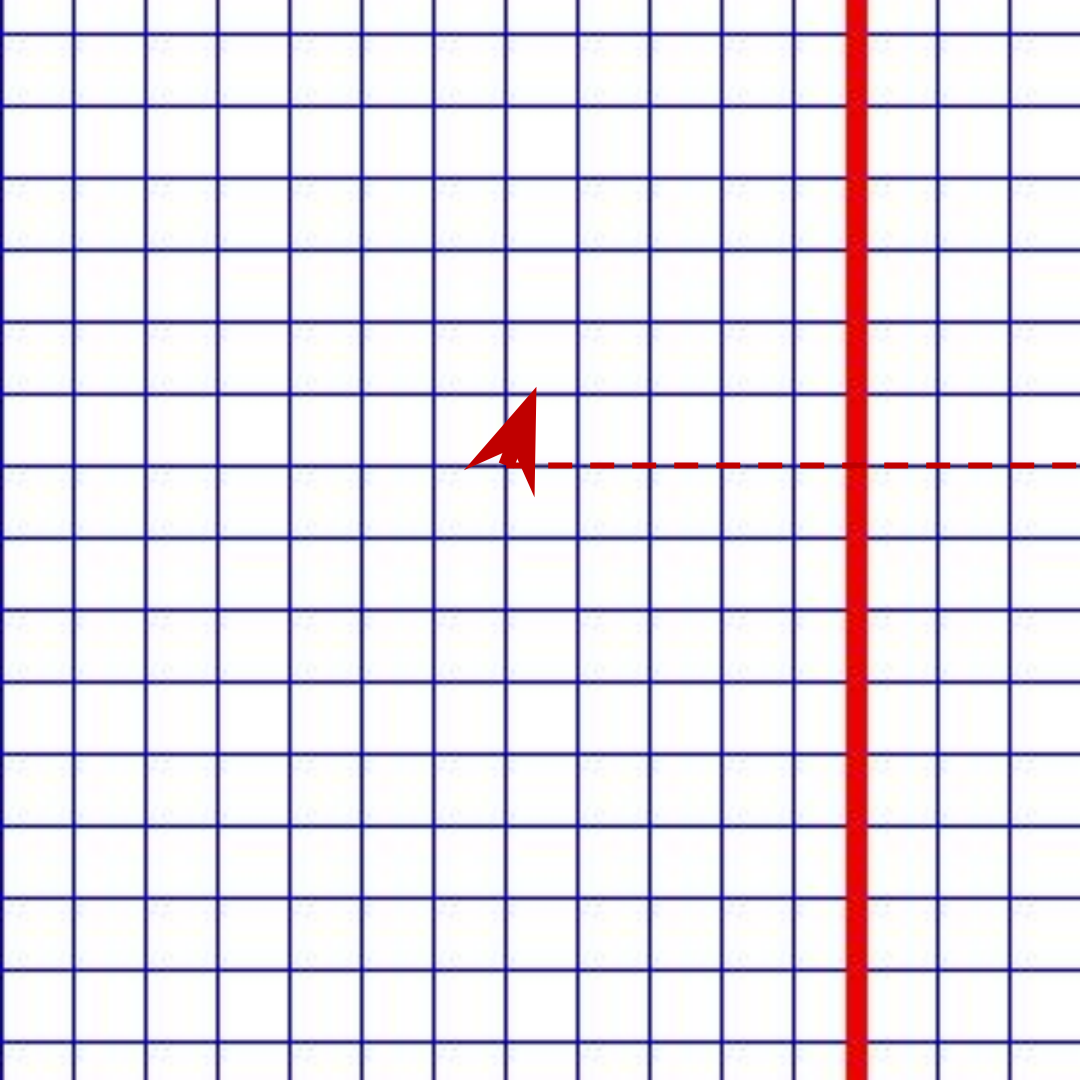
1. На сколько сдвигается по горизонтали и по вертикали.
2. Угол и длина



Способы описания вектора

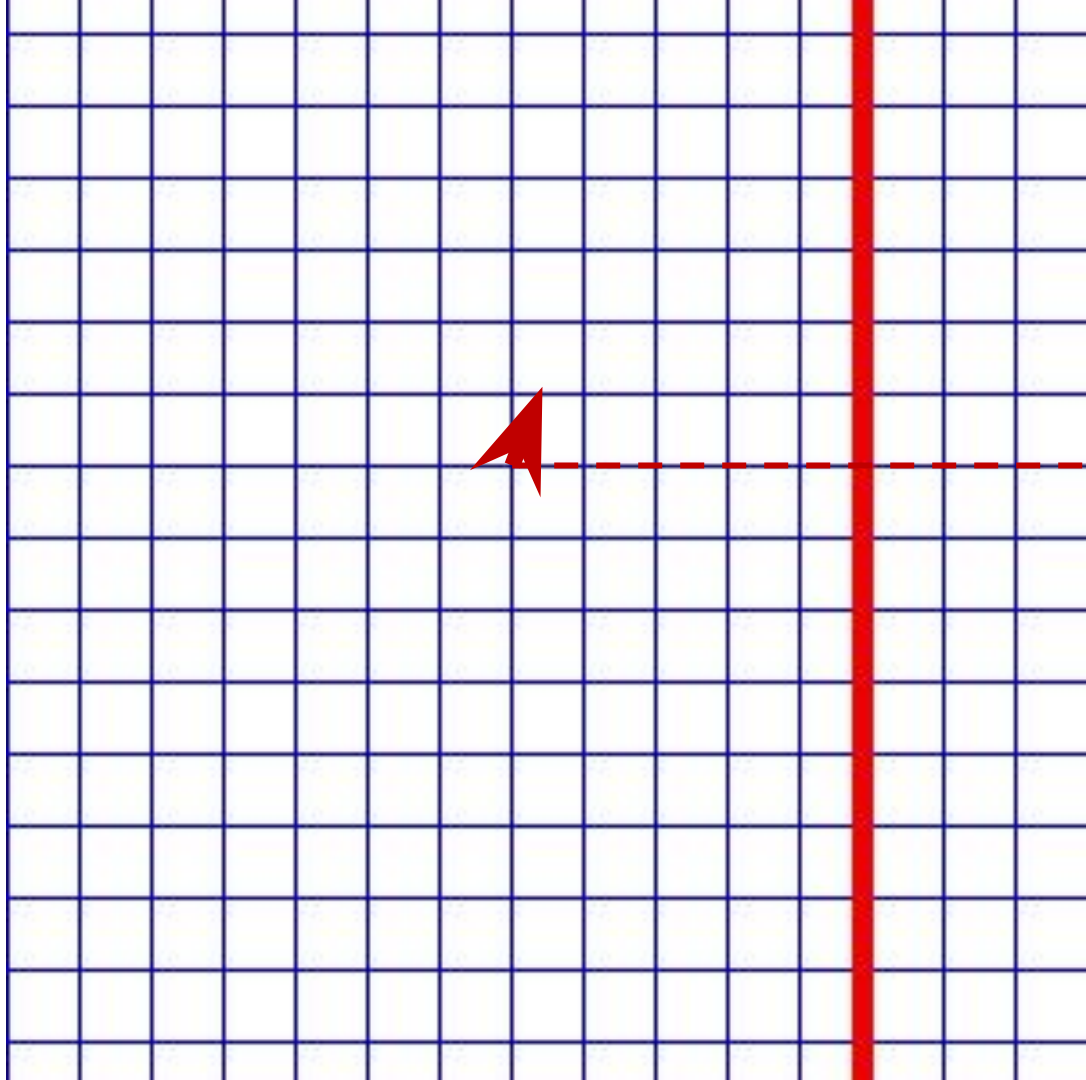
1. На сколько сдвигается по горизонтали и по вертикали.
2. Угол и длина

Тоже два числа. И меньше не получится, т.к. пространство двумерное.



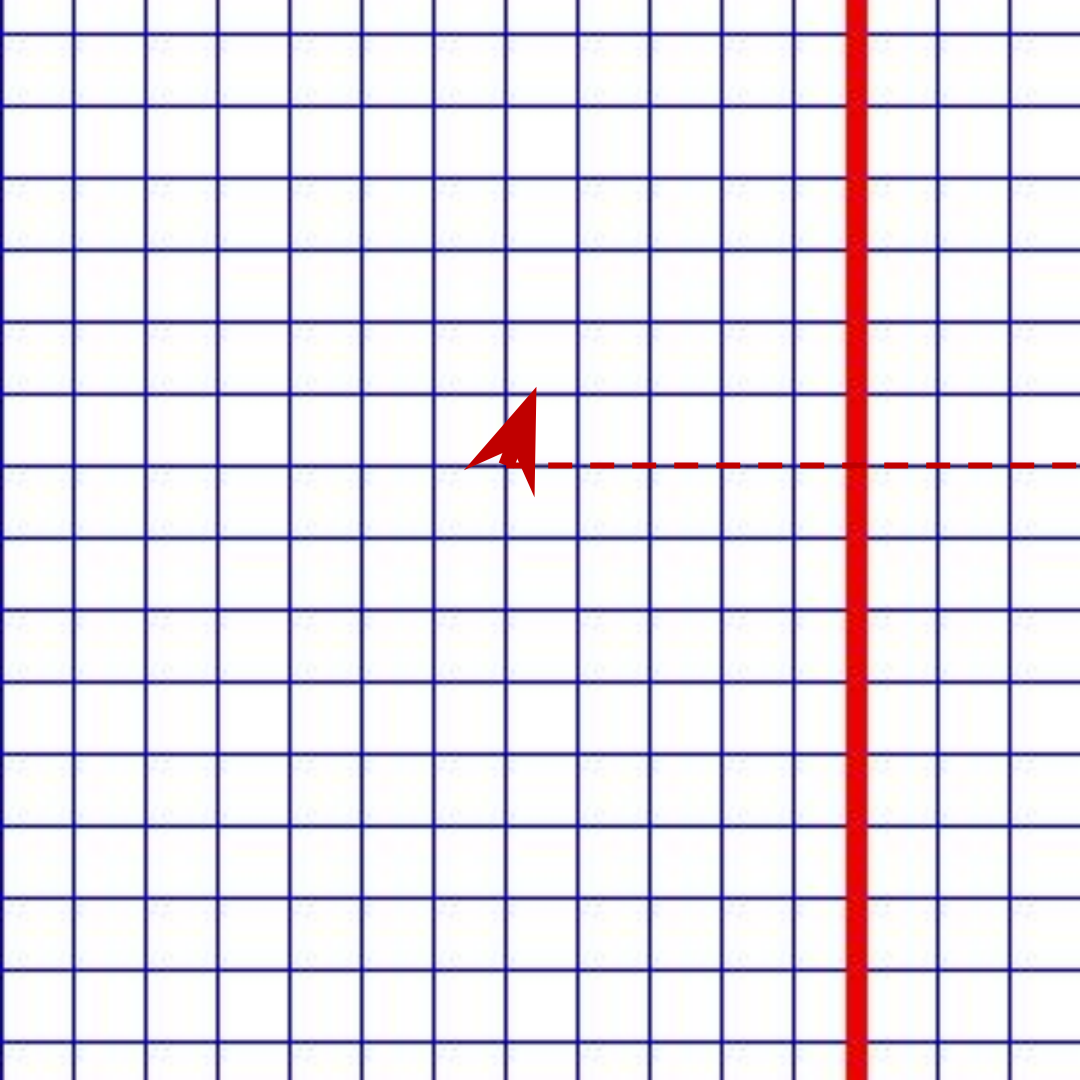
Способы описания вектора

1. *Декартова система координат.* На сколько сдвигается по горизонтали и по вертикали.
2. Угол и длина

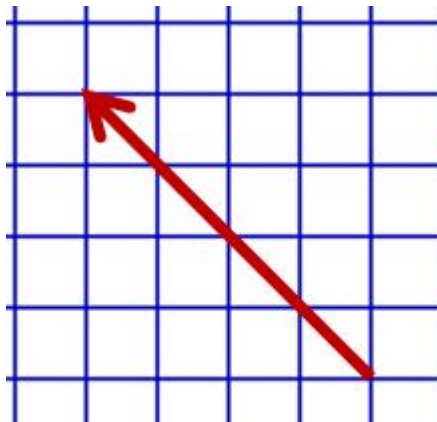


Способы описания вектора

1. *Декартова система координат.* На сколько сдвигается по горизонтали и по вертикали.
2. *Полярная система координат.* Угол и длина



Физика



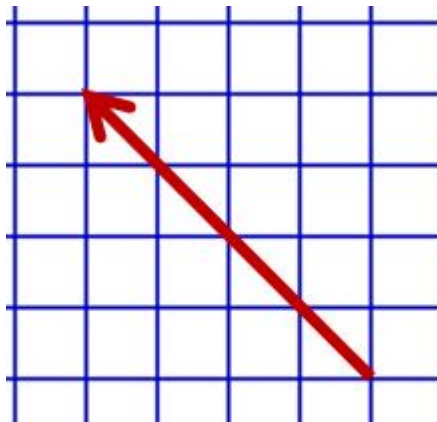
Math

?

DS

?

Физика



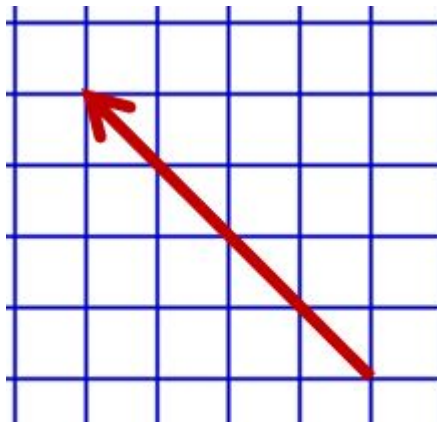
Math

$$\begin{bmatrix} -4 \\ 4 \end{bmatrix}$$

DS

?

Физика



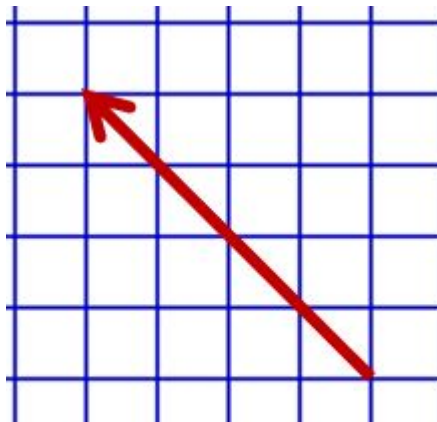
Math

$$\begin{bmatrix} -4 \\ 4 \end{bmatrix}$$

DS

Этаж	Площадь
-4	4

Физика



Math

$$\begin{bmatrix} -4 \\ 4 \end{bmatrix}$$

DS

Этаж	Площадь
-4	4



Операции

1. Сложение
2. Умножение на число

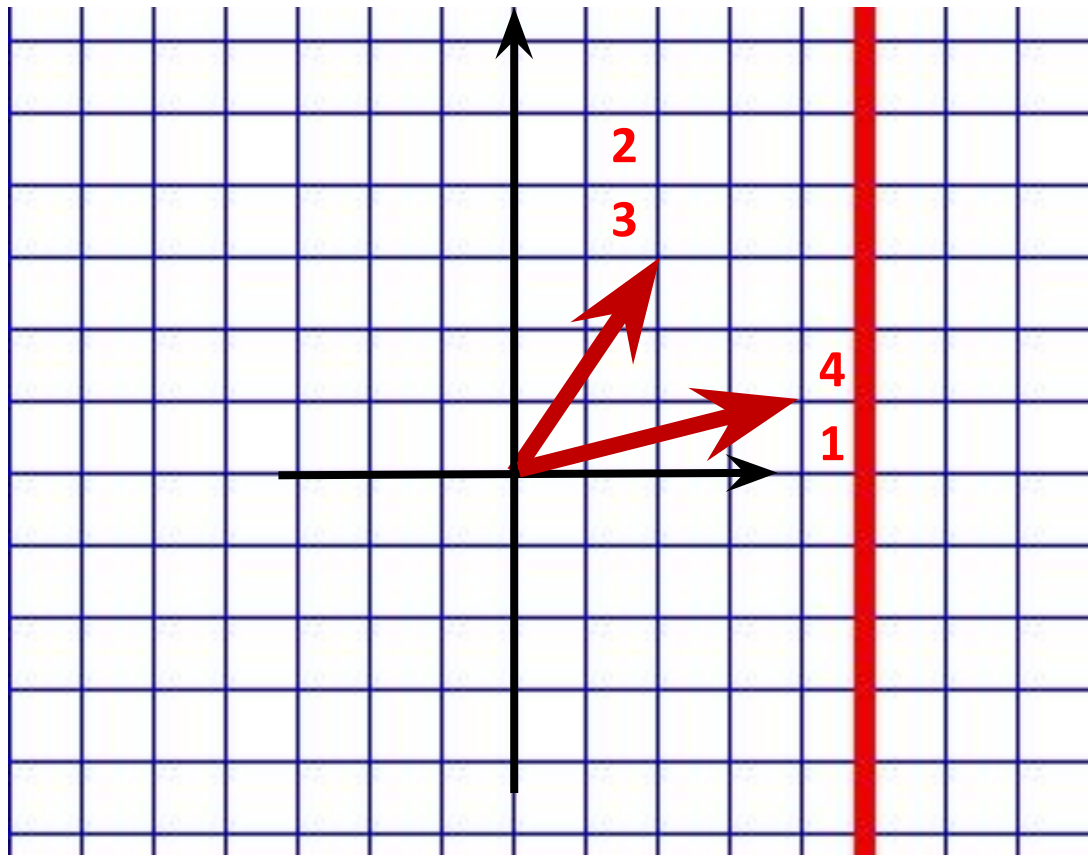
https://youtu.be/TJGH_5VMAL4

4:47 - 5:44

6:02 - 8:07

Операции

1. Сложение



Операции

1. Сложение
2. Умножение на число

А вычитание?

Операции

1. Сложение
2. Умножение на число

А вычитание?

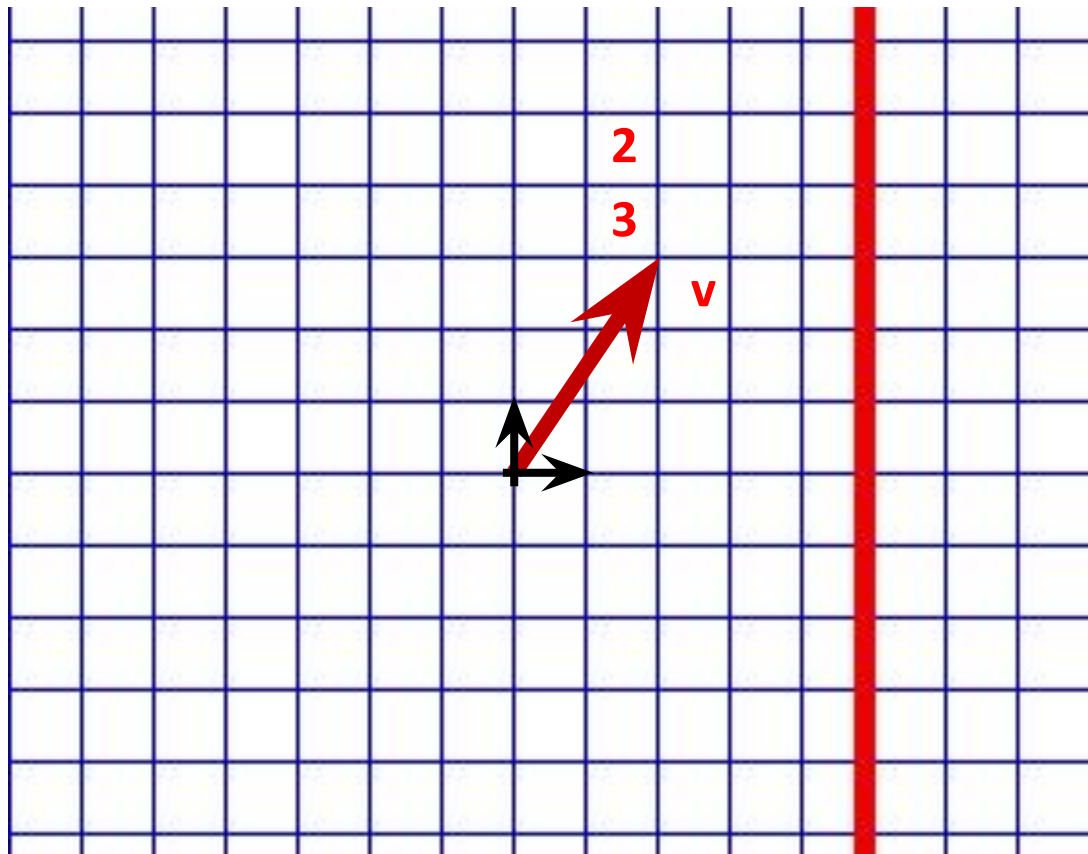
А как описать числами поворот?

Система координат

$$\bar{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\bar{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\bar{v} = 2\bar{i} + 3\bar{j}$$



Резюме

<https://youtu.be/gMsQWoVG3Ao>

0:50 - 3:45

Списки vs. Массивы

Структуры данных



Массив	Список
Фиксированного размера	
Один тип	
Непрерывная область памяти	

```
int Numbers[5];
```

```
numbers: List
```

```
String Names[2] = {"Anna", "Akakij"};
```

```
names = ["Anna", 5]
```

NumPy

Зачем?



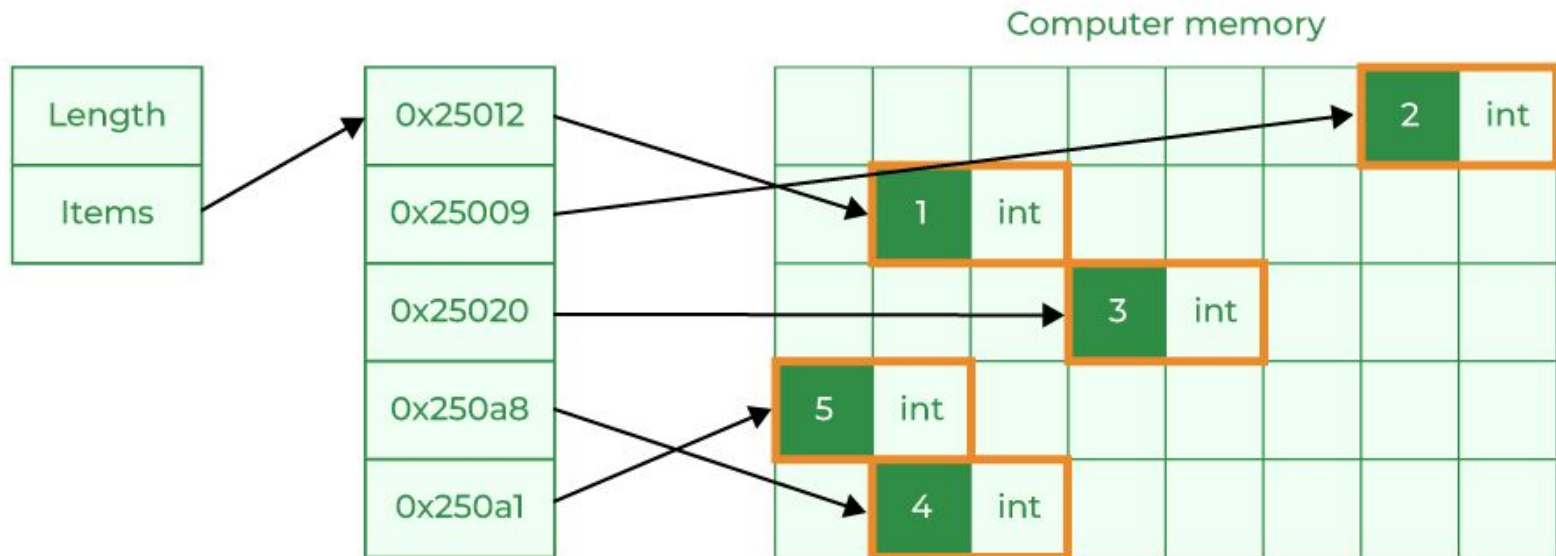
Массивы NumPy:

1. Быстрее и компактнее, чем списки Python
2. Потребляет меньше памяти

Хранение



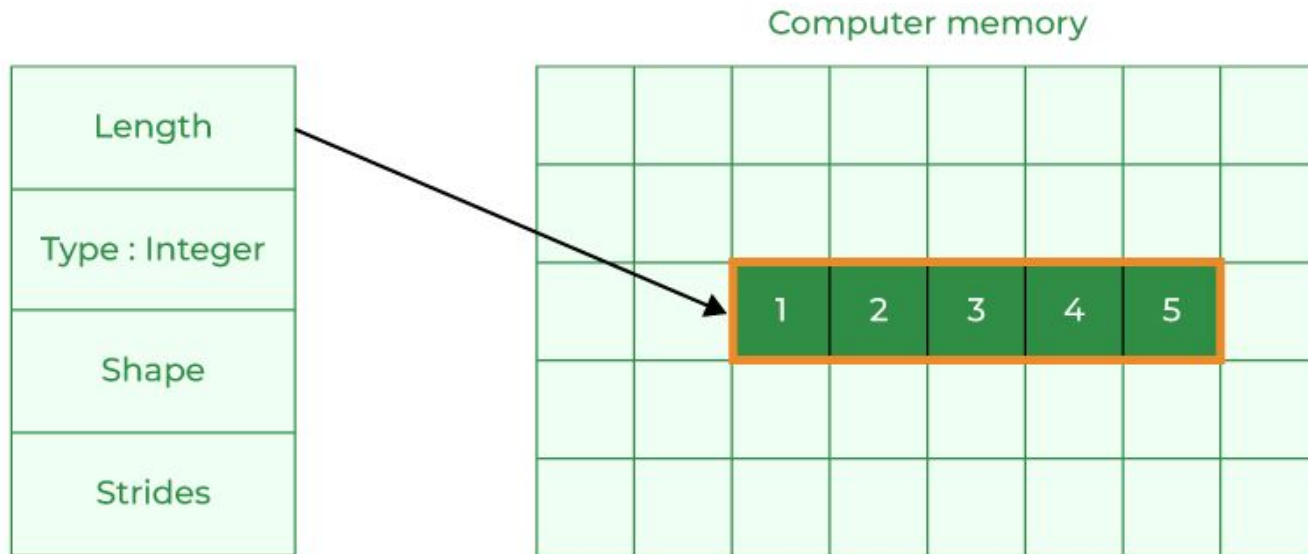
Python List



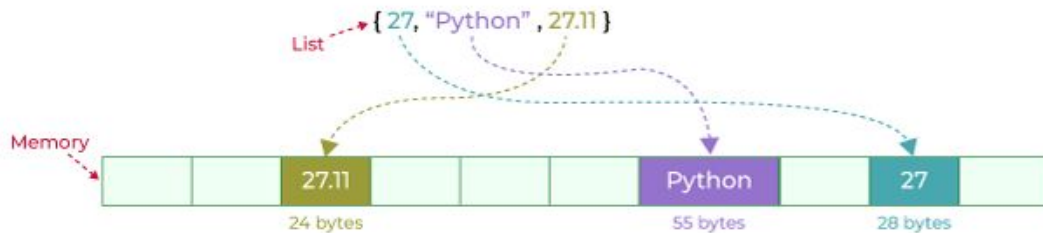
Хранение



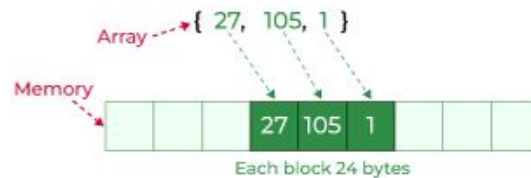
NumPy Array



Хранение



Python List



NumPy Array

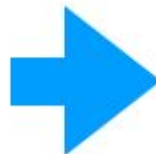
- Memory size in bytes are in Mac OS X.

Создаем вектор



Command

```
np.array([1,2,3])
```



NumPy Array

1
2
3

Готовые варианты



`np.ones(3)`



1
1
1

`np.zeros(3)`



0
0
0

`np.random.random(3)`



0.5967
0.0606
0.2223

Numpy.ndarray

`.shape` / `.ndim` / `.dtype` / `.itemsize` / `.size`

0	1	2	3	4
5	6	7	8	9
10	11	12	13	14

`.dtype`
= int32

`.itemsize`
= 4 (int32)

`.shape`
= (3, 5)

`.ndim`
= 2

`.size`
= 15

2

Готовые варианты



shape=(3, 4)

1	2	3	4
5	6	7	8
9	10	11	12

shape=(1,)

1	1	1	1
1	1	1	1
1	1	1	1

shape=(4,)

1	2	3	4
1	2	3	4
1	2	3	4

shape=(3,)

1	2	3	

shape=(3, 1)

1	1	1	1
5	5	5	5
9	9	9	9

Готовые варианты



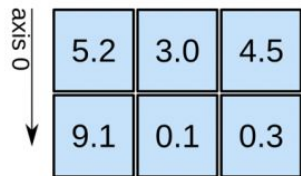
1D array



axis 0 →

shape: (4,)

2D array

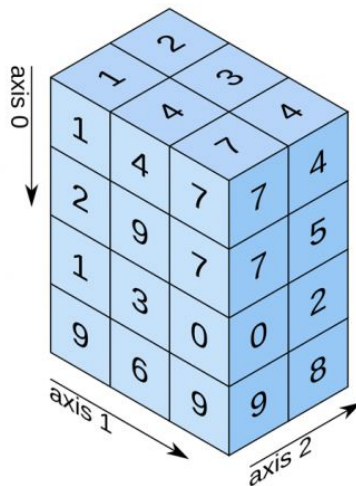


axis 0 ↓

axis 1 →

shape: (2, 3)

3D array



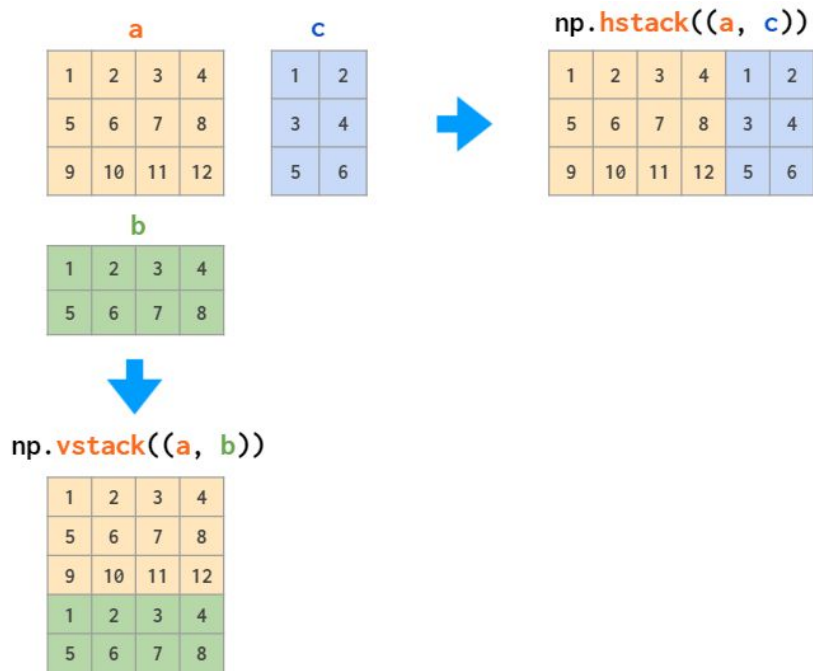
axis 0 ↓

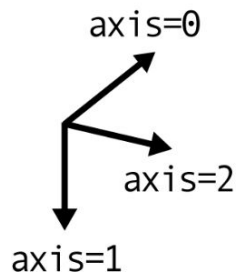
axis 1 →

axis 2 →

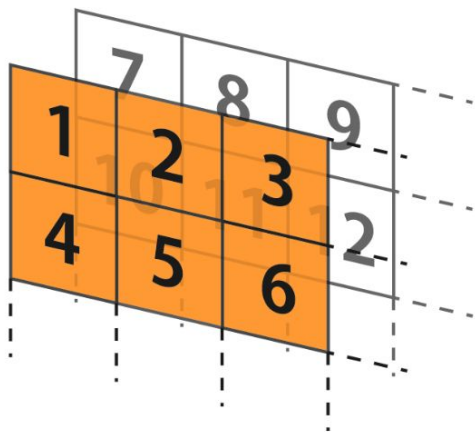
shape: (4, 3, 2)

Готовые варианты

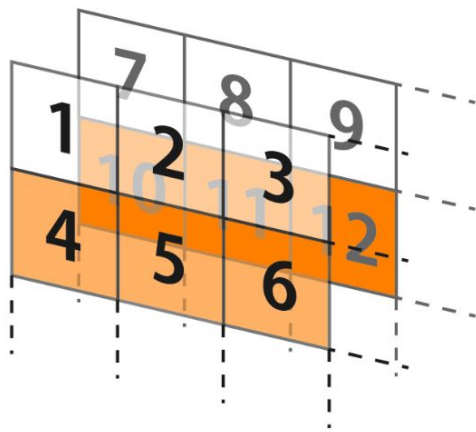




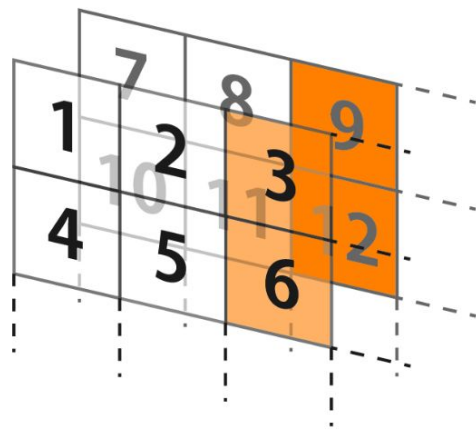
$X[0, :, :]$



$X[:, 1, :]$

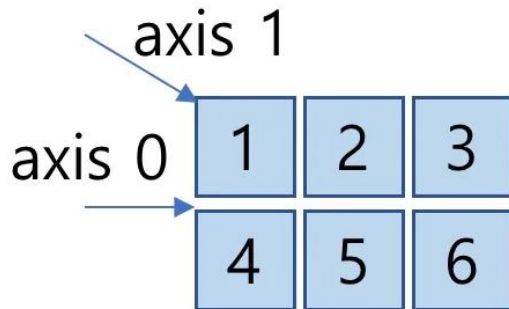


$X[:, :, 2]$

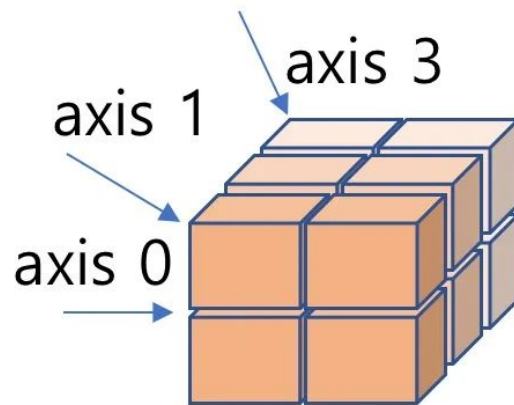




1D array



2D array



3D array

```
import numpy as np  
v = np.array([1, 2])
```

А вектор-строку как?

```
import numpy as np  
v = np.array([[1], [2]])
```

А вектор-строку как?

Итоги

Что было?



1. Важные свойства кодирования объектов
2. Вектор, эмбединг
3. NumPy – эффективно, компактно

<https://otus.ru/polls/122222/>

Дополнительные ссылки:

- <https://tproger.ru/articles/osnovy-raboty-s-massivami-i-spiskami>
- <https://habr.com/ru/articles/768188/> – **np.array** vs. **list**
- <https://www.geeksforgeeks.org/python-lists-vs-numpy-arrays/>
- [https://kvant.mccme.ru/1976/04/chto takoe vektor.htm](https://kvant.mccme.ru/1976/04/chto_takoe_vektor.htm)
- https://docs.google.com/presentation/d/186rD3ZEGuZOr8PnFr-Cxz1nG0IvZYhEr_od9FjKbhh0/edit?slide=id.ga3c3cc752b_0_280#slide=id.ga3c3cc752b_0_280
- <https://habr.com/ru/articles/728568/>