

ML Basic

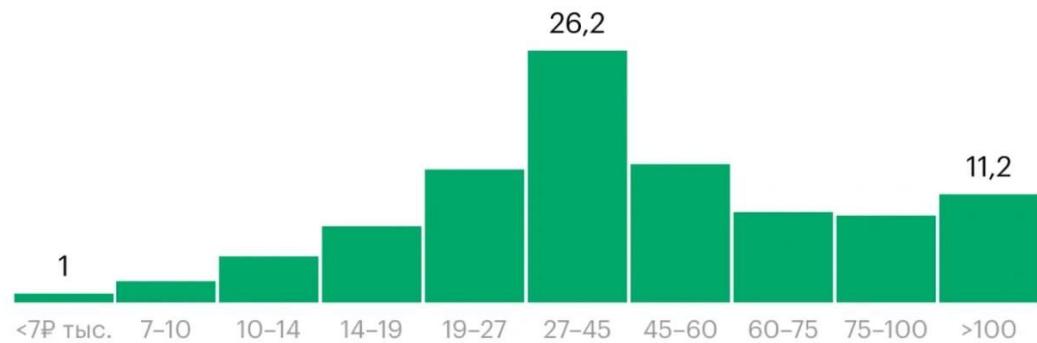
Визуализация

Начнем в 20:01 мск

ЧТО "НЕ ТАК"? ↗

Распределение населения России по величине среднедушевых доходов

Данные за 2023 год, %



Нижняя граница интервалов не включена

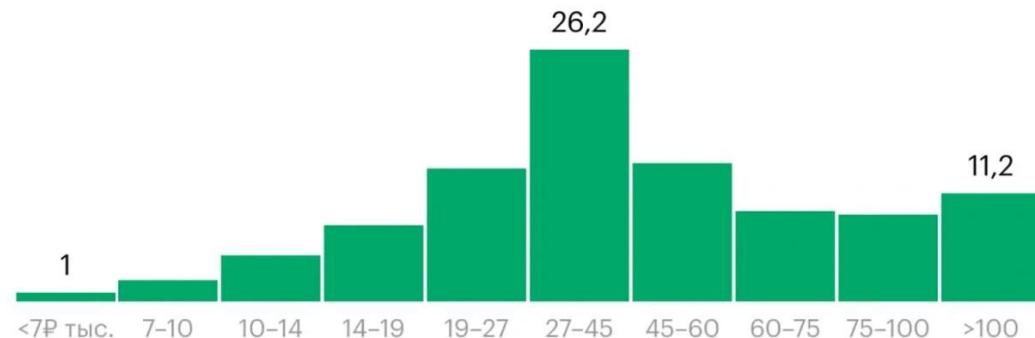


Warmup

Что “не так” с гистограммой?

Распределение населения России по величине среднедушевых доходов

Данные за 2023 год, %



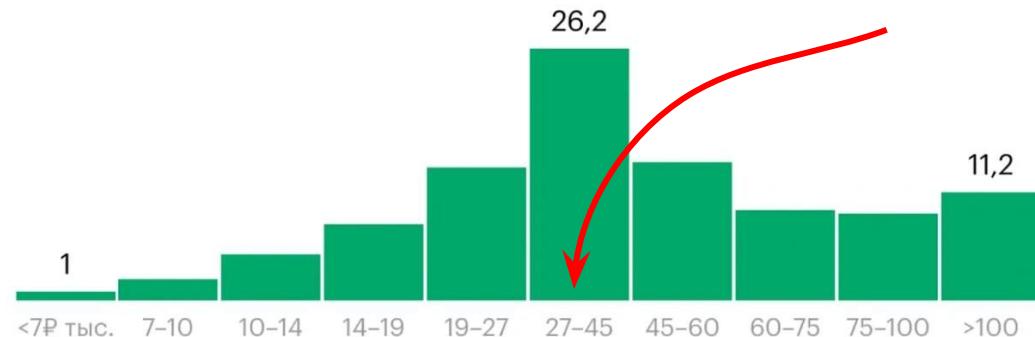
Нижняя граница интервалов не включена

Warmup

Что “не так” с гистограммой?

Распределение населения России по величине среднедушевых доходов

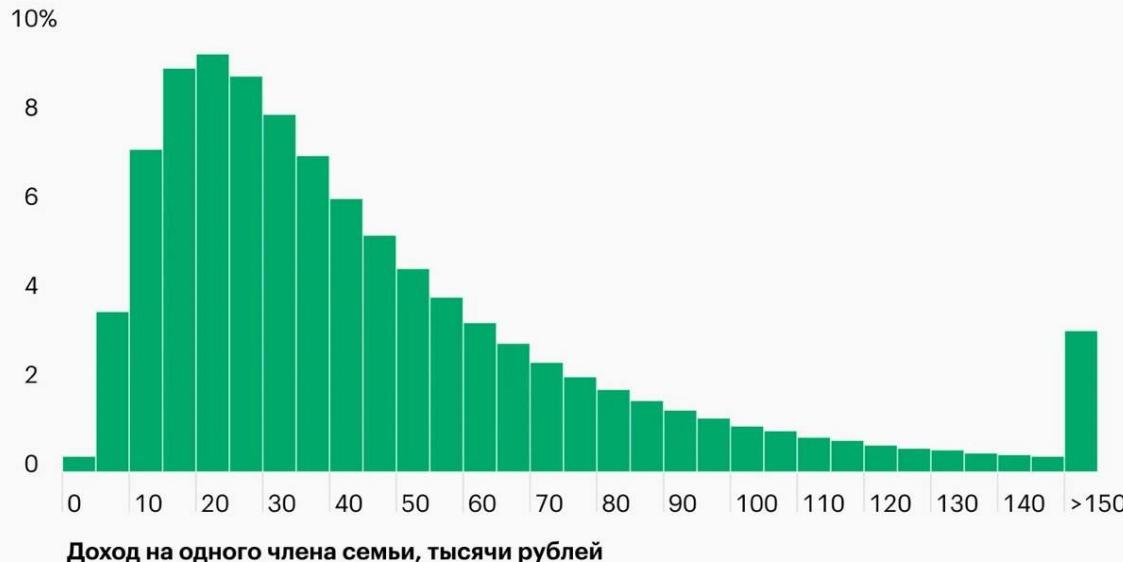
Данные за 2023 год, %



Нижняя граница интервалов не включена

Warmup

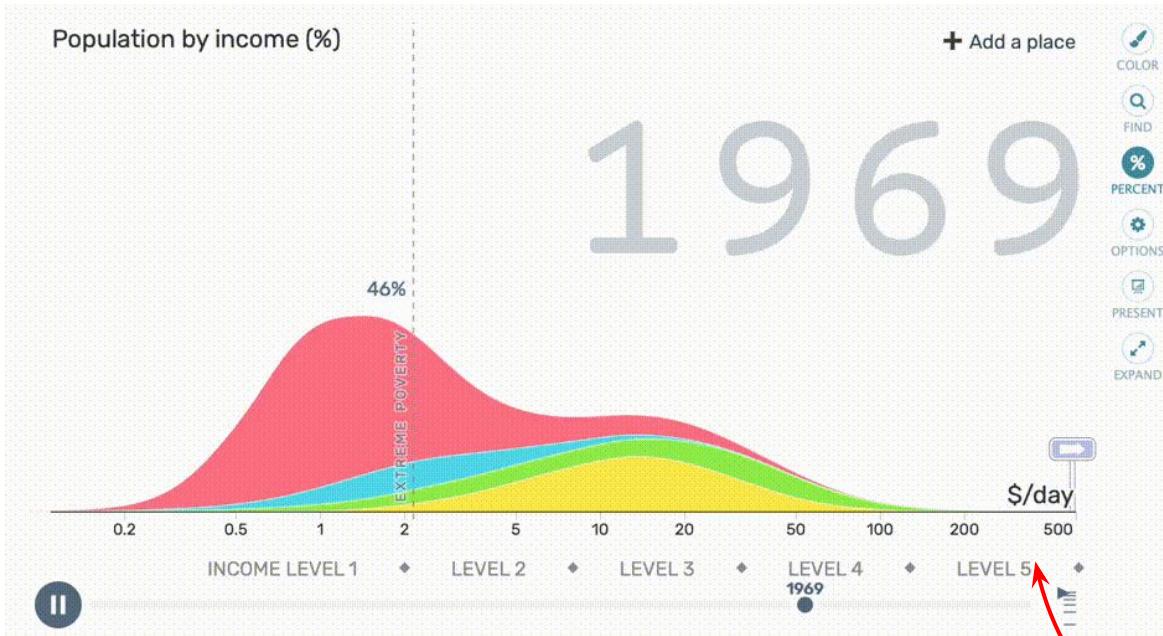
Распределение населения России по величине среднедушевых доходов в 2024 году



Источник:

<https://journal.tinkoff.ru/stat-manipulation-logical/>

Warmup



Источник:

[https://www.gapminder.org
/fw/income-mountains/](https://www.gapminder.org/fw/income-mountains/)

Половина людей в
мире живут на сумму
менее 8 долларов в
день

А тут тоже?

Нюанс 1: Паритет покупательской способности

How long do people need to work at minimum wage to buy one Big Mac?

2021 Big Mac price in US-Dollar according to the [Economist Big Mac Index](#) and 2019 minimum wage in US-Dollar in selected countries. The bigger a country circle, the bigger the population. [European countries are highlighted in blue](#).

Источник:

https://www.datawrapper.de/_UsFlq/

https://ru.wikipedia.org/wiki/Индекс_бигмака



Based on a chart by RJ Andrews.

Source: The Economist (Big Mac prices) and OECD (minimum wages) • Get the data • Created with Datawrapper

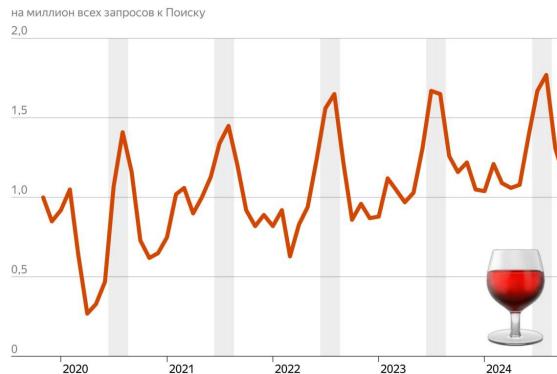
Где мы?

- **Dollar Street** – сайт где можно посмотреть как живут разные семьи в разных странах с определенным уровнем дохода,
- ... по многим параметрам – от домов и машин, до косметики, зубных щеток, зубов, состояния ванных комнат и тп.
- Проект делают уже 15 лет,
- 30 000 фотографий
- 264 семьи со всего мира
- Ссылка: <https://www.gapminder.org>

Рекомендации

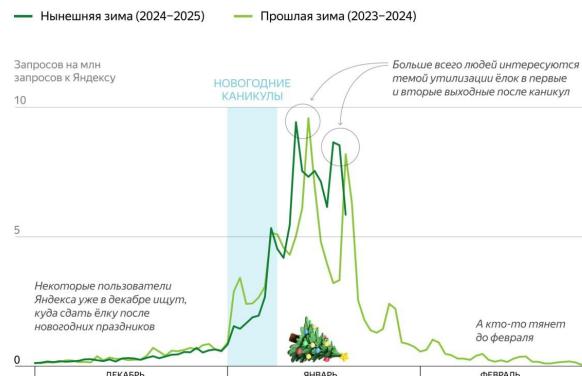
- http://t.me/data_csv – канал про data-журналистику
- <http://t.me/YaResearches> – канал от data-журналистов Яндекса

Количество запросов про дегустацию вина



ПО ДАННЫМ ПОИСКА ЯНДЕКСА, НОЯБРЬ 2019 — ОКТЯБРЬ 2024

Как меняется интерес к утилизации ёлок



Изучали запросы со словами сдать ёлку, приём ёлок, утилизация ёлок, переработка ёлок, выбросить ёлку.

ПО ДАННЫМ ПОИСКА ЯНДЕКСА

Рекомендации

- http://t.me/data_csv – канал с данными
- <http://t.me/YaResearches> – канал с исследованиями

Интерес к модным игрушкам в поиске Яндекса

Доля от всех запросов, %

0,08

0,06

0,04

0,02

0

Исследования Яндекса. По данным Поиска





• REC

Если запись не включена,
КРИЧИТЕ !

План

План



1. Ловушки визуализации и анализа
2. Основные модули для визуализации
 - Pandas `built-in`
 - Plotly
 - Matplotlib
 - ...

План



1. Ловушки визуализации и анализа
2. Основные модули для визуализации
 - Pandas `built-in`
 - Plotly
 - Matplotlib

Зачем?



- EDA = Exploratory Data Analysis = Разведочный анализ данных
- Мониторинг качества
- Помощь в принятии решения

Зачем?



- EDA = Exploratory Data Analysis = Разведочный анализ данных
- Мониторинг качества
- Помощь в принятии решения

Постановка задачи

Этапы обучения

1. Освоение
2. **Осмысление**
3. Практика

Мируквите Борука, 107 лет



Код формы по ОКУД _____
Код учреждения по ОКПО _____
Министерство здравоохранения
СССР
(наименование учреждения)
Лаборатория _____

Медицинская документация
Форма № 216/у
Утверждена Минздравом СССР
04.10.80 г. № 1030

АНАЛИЗ МОКРОТЫ № _____
« _____ » 20 _____ г. час. _____ мин.
(дата взятия биоматериала) Возраст _____
Фамилия, и., о. *Мируквите Борука*

Учреждение _____ отделение _____
Палата _____ участок _____ медицинская карта № _____
Соленость слезы _____ Запах _____
Цвет _____ Характер *вредный*
Температура _____
Пупырки *51 не обнаружены*

Все измерения производятся в условных единицах измерения, одинаковых для соответствующих значений



Кейс: Решение должно быть принято

Drive

Search in Drive

+ New

... > 02-1: Set description > demo ▾

Type ▾ People ▾ Modified ▾ Source ▾

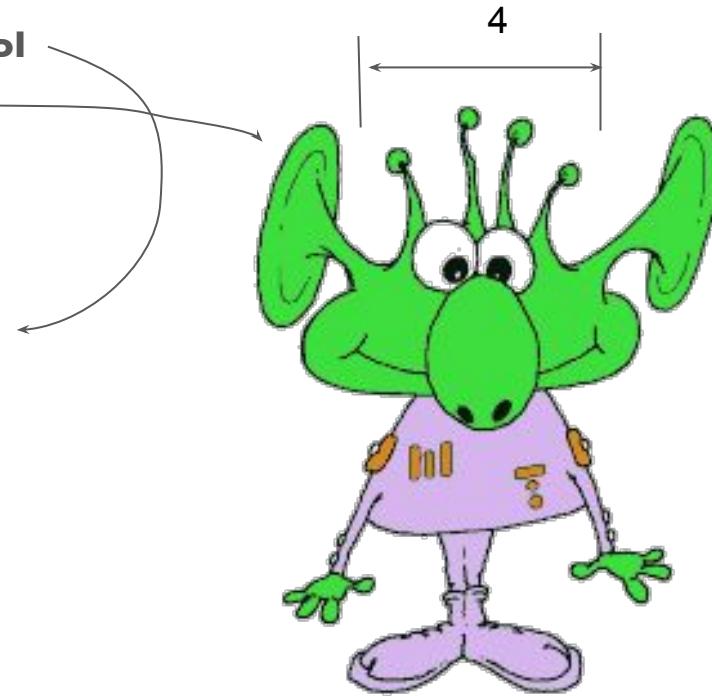
Name	Owner
fars_deseases	me



Кейс: Может бы так?

1. Научиться делать **сводные таблицы**
2. Придумать “хороший” **показатель**

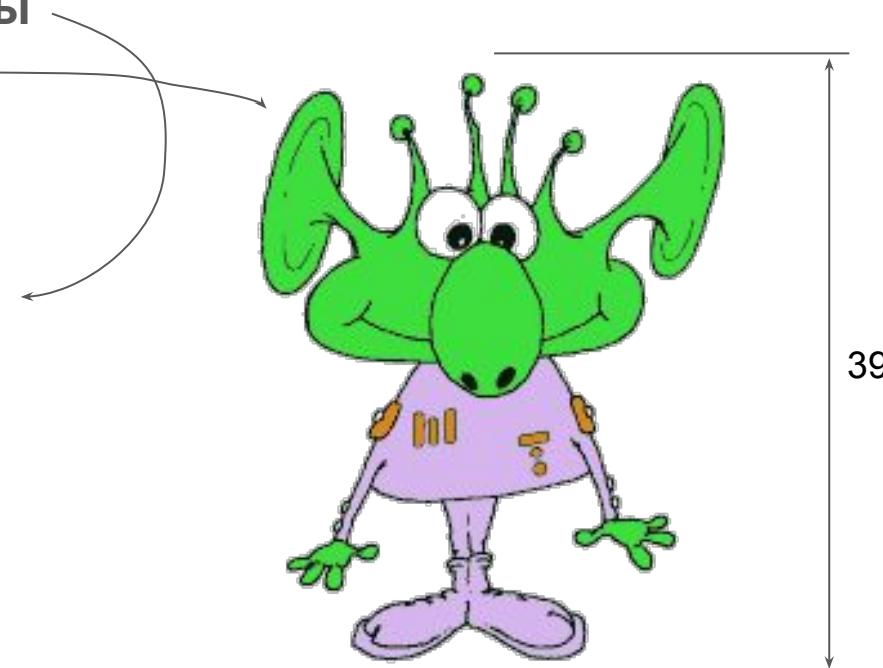
Диагноз	Хороший показатель
Здоровый	1
Оптимизм 1 степени	7
Оптимизм 2 степени	-5



Кейс: Лучше диапазон

1. Научиться делать **сводные таблицы**
2. Придумать “хороший” **показатель**

Диагноз	Хороший показатель
Здоровый	1 – 3
Оптимизм 1 степени	7 – ...
Оптимизм 2 степени	-... – -5





Кейс: Пример

Клинический анализ крови

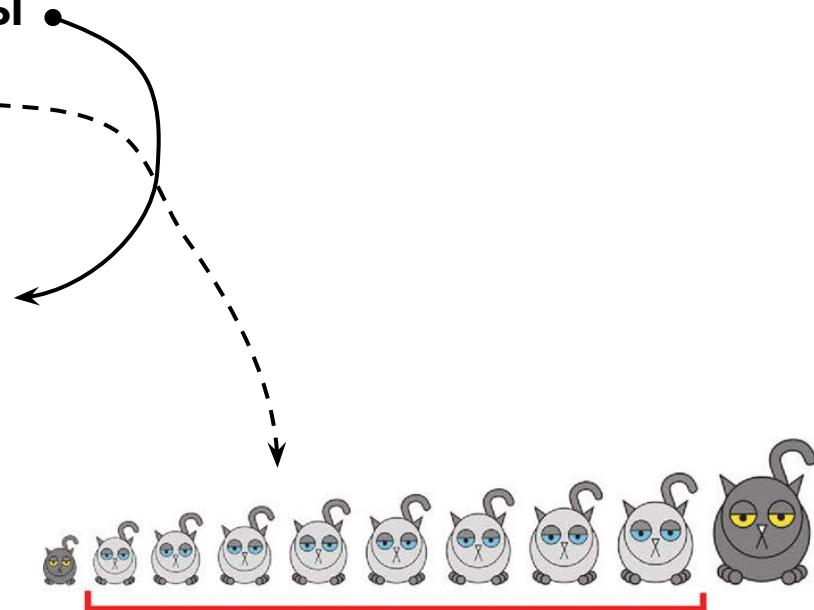
Исследование	Результат	Единицы	Референсные значения
Гематокрит	38.6*	%	39.0 - 49.0
Гемоглобин	13.8	г/дл	13.2 - 17.3
Эритроциты	4.60	млн/мкл	4.30 - 5.70
MCV (ср. объем эритр.)	83.9	фл	80.0 - 99.0
RDW (ширина распределения эритр.)	12.1	%	11.6 - 14.8
MCH (ср. содер. Hb в эр.)	30.0	пг	27.0 - 34.0
MCHC (ср. конц. Hb в эр.)	35.8	г/дл	32.0 - 37.0
Тромбоциты	159	тыс/мкл	150 - 400
Лейкоциты	4.69	тыс/мкл	4.50 - 11.00
Нейтрофилы (общ.число), %	42.5*	%	48.0 - 78.0



Кейс: Цель

1. Научиться делать **сводные таблицы**
2. Придумать “хороший” **показатель**

Диагноз	Среднее давление
Здоровый	1
Оптимизм 1 степени	70000
Оптимизм 2 степени	-50000



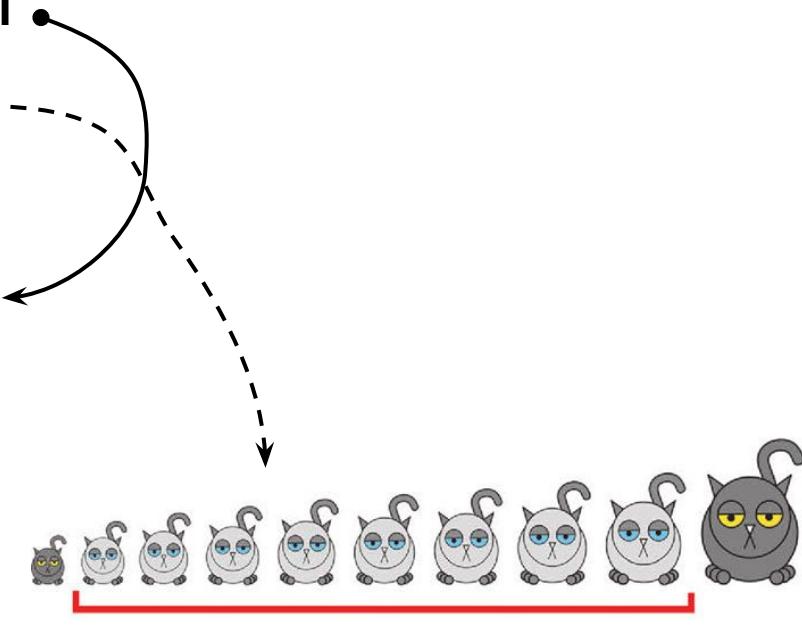
Котики для усеченного среднего



Кейс: Цель

1. Научиться делать **сводные таблицы**
2. Понять, как выбирать **интервал**

Диагноз	Среднее давление
Здоровый	1 – 3
Оптимизм 1 степени	7 – ...
Оптимизм 2 степени	-... – -5





Кейс: Как получить доступ?

The screenshot shows a spreadsheet application window titled "fars_deseases". The menu bar includes File, Edit, View, Insert, Format, Data, Tools, Extensions, and Help. The "File" menu is open, showing options like Import and Make a copy. A secondary context menu is open under the "Share" option, listing Share with others and Publish to web.

			G		
1			пушка наличие пупырь дл		
2	E		5	TRUE	
3	E		7	TRUE	
4	E	53.3	74	6	TRUE
5	E	54.3	88	4	TRUE
6	E	58.8	96	2	FALSE
7	E	53	88	2	TRUE
8	E	55	67	0	FALSE

Группировки

Фамилия	Район проживания	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	Адок	137	7	Оптимизм 2 степени
Бихошюев	Адок	72	7	Оптимизм 2 степени
Бидетаин	Хоккинг-стрит	88	5	Повышенная пупырчатость
Бицевюев	Хоккинг-стрит	90	5	Здоров
Бицедиова	Хоккинг-стрит	10	4	Оптимизм 1 степени
Бацегаин	Площадь Гагарина	20	4	Трихотомия
Бачытиин	Площадь Гагарина	60	4	Повышенная пупырчатость
Бикэнаова	Площадь Гагарина	67	0	Оптимизм 1 степени

```
data [data [ "Количество ушей" ] ==4 ] [ "Сolenость слезы" ].mean ()
```

Фамилия	Район проживания	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	Адок	137	7	Оптимизм 2 степени
Бихошюев	Адок	72	7	Оптимизм 2 степени
Бидетаин	Хоккинг-стрит	88	5	Повышенная пупырчатость
Бицевюев	Хоккинг-стрит	90	5	Здоров
Бицедиова	Хоккинг-стрит	10	4	Оптимизм 1 степени
Бацегаин	Площадь Гагарина	20	4	Трихотомия
Бачытиин	Площадь Гагарина	60	4	Повышенная пупырчатость
Бикэнаова	Площадь Гагарина	67	0	Оптимизм 1 степени

```
data[data["Количество ушей"]==4] ["Сolenость слезы"].max()
```



Фильтр



Отбор столбцов



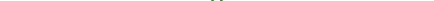
Агрегация

Фамилия	Район проживания	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	Адок	137	7	Оптимизм 2 степени
Бихошюев	Адок	72	7	Оптимизм 2 степени
Бидетаин	Хоккинг-стрит	88	5	Повышенная пупырчатость
Бицевюев	Хоккинг-стрит	90	5	Здоров
Бицедиова	Хоккинг-стрит	10	4	Оптимизм 1 степени
Бацегаин	Площадь Гагарина	20	4	Трихотомия
Бачытиин	Площадь Гагарина	60	4	Повышенная пупырчатость
Бикэнарова	Площадь Гагарина	67	0	Оптимизм 1 степени

```
data[data["Количество ушей"]==4]["Сolenость слезы"].max()
```

Curved red line with a small gap at the bottom right corner.

Фильтр

✓  

Отбор столбцов

Агрегация

Фамилия	Район проживания	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	Адок	137	7	Оптимизм 2 степени
Бихошюев	Адок	72	7	Оптимизм 2 степени
Бидетаин	Хоккинг-стрит	88	5	Повышенная пупырчатость
Бицевюев	Хоккинг-стрит	90	5	Здоров
Бицедиова	Хоккинг-стрит	10	4	Оптимизм 1 степени
Бацегаин	Площадь Гагарина	20	4	Трихотомия
Бачытиин	Площадь Гагарина	60	4	Повышенная пупырчатость
Бикэнаова	Площадь Гагарина	67	0	Оптимизм 1 степени

```
data[data[ "Количество ушей" ]==4] [ "Сolenость слезы" ].max()
```

Фильтр

Отбор столбцов

Агрегация

Фамилия	Район проживания	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	Адок	137	7	Оптимизм 2 степени
Бихошюев	Адок	72	7	Оптимизм 2 степени
Бидетаин	Хоккинг-стрит	88	5	Повышенная пупырчатость
Бицевюев	Хоккинг-стрит	90	5	Здоров
Бицедиова	Хоккинг-стрит	10	4	Оптимизм 1 степени
Бацегаин	Площадь Гагарина	20	4	Трихотомия
Бачытиин	Площадь Гагарина	60	4	Повышенная пупырчатость
Бикэнаова	Площадь Гагарина	67	0	Оптимизм 1 степени

```
data[data["Количество ушей"]==4]["Сolenость слезы"].max()
```



Фильтр



Отбор столбцов



Агрегация

Фамилия	Район проживания	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	Адок	137	7	Оптимизм 2 степени
Бихошюев	Адок	72	7	Оптимизм 2 степени
Бидетаин	Хоккинг-стрит	88	5	Повышенная пупырчатость
Бицевюев	Хоккинг-стрит	90	5	Здоров
Бицедиова	Хоккинг-стрит	10	4	Оптимизм 1 степени
Бацегаин	Площадь Гагарина	20	4	Трихотомия
Бачытиин	Площадь Гагарина	60	4	Повышенная пупырчатость
Бикэнаова	Площадь Гагарина	67	0	Оптимизм 1 степени

```
data[data["Количество ушей"] == 4] ["Сolenость слезы"].min()
```

Фильтр

Отбор столбцов

Агрегация

Фамилия	Район проживания	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	Адок	137	7	Оптимизм 2 степени
Бихошюев	Адок	72	7	Оптимизм 2 степени
Бидетаин	Хоккинг-стрит	88	5	Повышенная пупырчатость
Бицевюев	Хоккинг-стрит	90	5	Здоров
Бицедиова	Хоккинг-стрит	10	4	Оптимизм 1 степени
Бацегаин	Площадь Гагарина	20	4	Трихотомия
Бачытиин	Площадь Гагарина	60	4	Повышенная пупырчатость
Бикэнаова	Площадь Гагарина	67	0	Оптимизм 1 степени

```
data[data["Количество ушей"] == 4] ["Сolenость слезы"].count()
```



Фильтр



Отбор столбцов



Агрегация

Фамилия	Район проживания	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	Адок	137	7	Оптимизм 2 степени
Бихошюев	Адок	72	7	Оптимизм 2 степени
Бидетаин	Хоккинг-стрит	88	5	Повышенная пупырчатость
Бицевюев	Хоккинг-стрит	90	5	Здоров
Бицедиова	Хоккинг-стрит	10	4	Оптимизм 1 степени
Бацегаин	Площадь Гагарина	20	4	Трихотомия
Бачытиин	Площадь Гагарина	60	4	Повышенная пупырчатость
Бикэнаова	Площадь Гагарина	67	0	Оптимизм 1 степени

```
data[data["Количество ушей"] == 4] ["Сolenость слезы"].mean()
```



Фильтр



Отбор столбцов



Агрегация

Фамилия	Соленость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	max (Соленость слезы)
Оптимизм 2 степени	137
Пузырчатость	90
Здоров	67

Хотим так ...

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	max (Сolenость слезы)
Оптимизм 2 степени	137

А, если только для “Оптимизма 2 степени”?

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	max (Сolenость слезы)
Оптимизм 2 степени	137

А, если только для “Оптимизма 2 степени”?

```
data[data[ "Диагноз" ] == "Оптимизм 2 степени"]
```

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	max (Сolenость слезы)
Оптимизм 2 степени	137

А, если только для “Оптимизма 2 степени”?

```
data[data[ "Диагноз" ] == "Оптимизм 2 степени"] [ "Сolenость" ]
```

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	max (Сolenость слезы)
Оптимизм 2 степени	137

А, если только для “Оптимизма 2 степени”?

```
data[data[ "Диагноз" ] == "Оптимизм 2 степени"] [ "Сolenость" ].max()
```

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	max (Сolenость слезы)
Оптимизм 2 степени	137
Пузырчатость	90
Здоров	67

Теперь для всех!

```
data[data["Диагноз"] == "Оптимизм 2 степени"]["Сolenость"].max()
```

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	max (Сolenость слезы)
Оптимизм 2 степени	137
Пузырчатость	90
Здоров	67

Теперь для всех!

```
data[data["Диагноз"] == "Оптимизм 2 степени"]["Сolenость"].max()
```

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пупырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	max (Сolenость слезы)
Оптимизм 2 степени	137
Пупырчатость	90
Здоров	67

Теперь для всех!

```
data.groupby(by="Диагноз")["Сolenость"].max()
```

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пупырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	
Оптимизм 2 степени	<u>137</u>
Пупырчатость	<u>90</u>
Здоров	<u>67</u>

`data.groupby(by='Диагноз') ['Сolenость слезы'].max()`

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пупырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	
Оптимизм 2 степени	<u>137</u>
Пупырчатость	<u>90</u>
Здоров	<u>67</u>

`data.groupby(by='Диагноз') ['Сolenость слезы'].max()`

`data[data['Диагноз'] == 'Пупырчатость'] ['Сolenость слезы'].max()`

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнаова	67	0	Здоров

Диагноз	
Оптимизм 2 степени	
Пузырчатость	
Здоров	

```
data.groupby(by='Диагноз')[['Сolenость слезы']].min()
```

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	
Оптимизм 2 степени	<u>88</u>
Пузырчатость	<u>90</u>
Здоров	<u>10</u>

`data.groupby(by='Диагноз') ['Сolenость слезы'].min()`

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнаова	67	0	Здоров

Диагноз	
Оптимизм 2 степени	
Пузырчатость	
Здоров	

`data.groupby(by='Диагноз') ['Количество ушей'].min()`

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнаова	67	0	Здоров

Диагноз	
Оптимизм 2 степени	5
Пузырчатость	5
Здоров	0

`data.groupby(by='Диагноз') ['Количество ушей'].min()`

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	

`data.groupby(by='Количество ушей')['Диагноз'].min()`

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пупырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

7	
5	
4	
0	

`data.groupby(by='Количество ушей')['Диагноз'].min()`

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	
0	Здоров
4	Здоров
5	Оптимизм
7	Оптимизм

```
data.groupby(by='Количество ушей')[ 'Диагноз' ].min()
```

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пупырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнаова	67	0	Здоров

Диагноз	
7	
5	
4	
0	

```
data.groupby(by='Количество ушей')[ 'Диагноз' ].count()
```

Фамилия	Сolenость слезы	Количество ушей	Диагноз
Бихивеин	137	7	Оптимизм 2 степени
Бихошюев	72	7	Оптимизм 2 степени
Бидетаин	88	5	Оптимизм 2 степени
Бицевюев	90	5	Пузырчатость
Бицедиова	10	4	Здоров
Бацегаин	20	4	Здоров
Бачытиин	60	4	Здоров
Бикэнарова	67	0	Здоров

Диагноз	
7	2
5	2
4	3
0	1

```
data.groupby(by='Количество ушей')[ 'Диагноз' ].count()
```

Что получилось?



Кейс: Итог

```
▶ data.groupby(by='Диагноз')[['Соленость слезы']].mean()
```



Соленость слезы

Диагноз

Грибб	103.673469
Дихотомия	101.566667
Здоров	100.287129
Оптимизм 1 степени	104.160000
Оптимизм 2 степени	102.820000
Повышенная пупырчатость	99.700000
Трихотомия	99.510204

dtype: float64



Кейс: Визуализация

```
▶ data.groupby(by='Диагноз')['Соленость слезы'].mean()
```



Соленость слезы

Диагноз

Грибб	103.673469
Дихотомия	101.566667
Здоров	100.287129
Оптимизм 1 степени	104.160000
Оптимизм 2 степени	102.820000
Повышенная пупырчатость	99.700000
Трихотомия	99.510204

dtype: float64

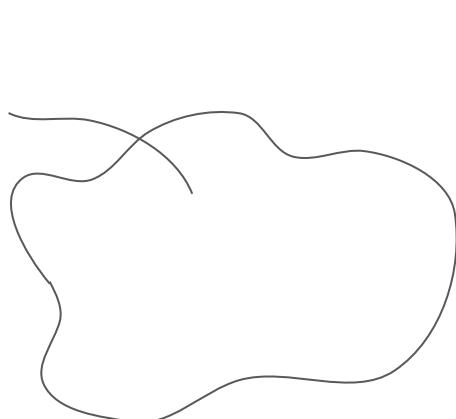
Характеристика множества



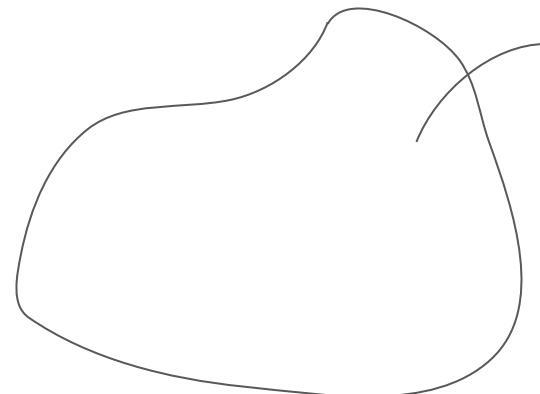
Характеристика множества

- Кто наши клиенты?
- В каком городе открывать филиал?
- Какая аудитория более лояльная?

42



100 500



Среднее арифметическое



Среднее арифметическое. Плюсы

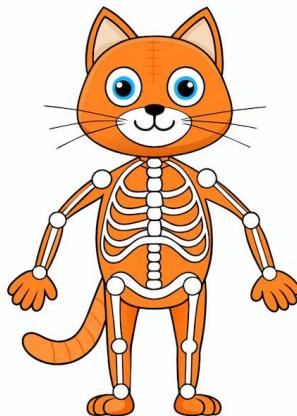
1. Общеизвестно
2. Легко вычислить
3. Дифференцируемо

$$\frac{x_1 + x_2 + \dots + x_n}{n}$$

1



Среднее арифметическое. Минусы



Среднее
число скелетов
в организме кота:

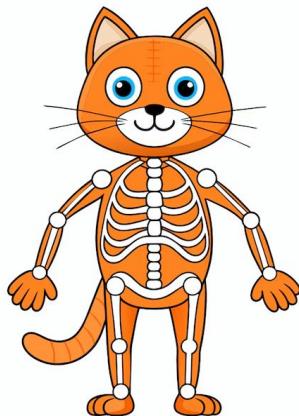
A: Меньше 1

B: Равно 1

C: Превышает 1



Среднее арифметическое. Минусы



Среднее
число скелетов
в организме кота:

A: Меньше 1

B: Равно 1

C: Превышает 1



Среднее арифметическое. Минусы

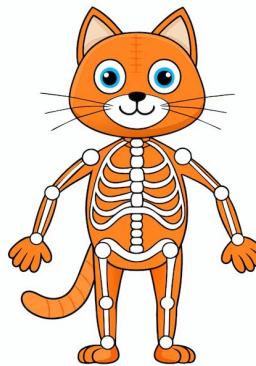


Среднее
число скелетов
в организме кота превышает 1

$$\frac{x_1 + x_2 + \dots + x_n}{n} =$$



Среднее арифметическое. Минусы



Среднее
число скелетов
в организме кота превышает 1

$$\frac{x_1 + x_2 + \dots + x_n}{n}$$

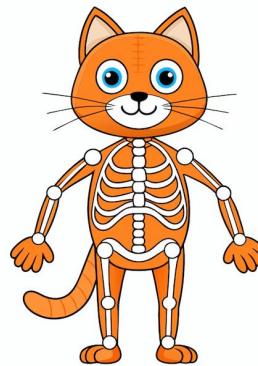
=



5



Среднее арифметическое. Минусы



Среднее
число скелетов
в организме кота превышает 1

$$\frac{x_1 + x_2 + \dots + x_n}{n} =$$



5



Среднее арифметическое. Минусы



Среднее
число скелетов
в организме кота превышает 1

$$\frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\text{_____}}{12}$$

=





Среднее арифметическое. Минусы

1. Может трактоваться неоднозначно

2



Среднее арифметическое. Минусы

1. Может трактоваться неоднозначно
2. Плохой показатель типичности

 ! Предупреждение: у нас есть цензура и предварительный
покиньте сайт. 18+

**Анекдоты
из России**
выходят с 8 ноября 1995 года
составитель - Дима Вернер

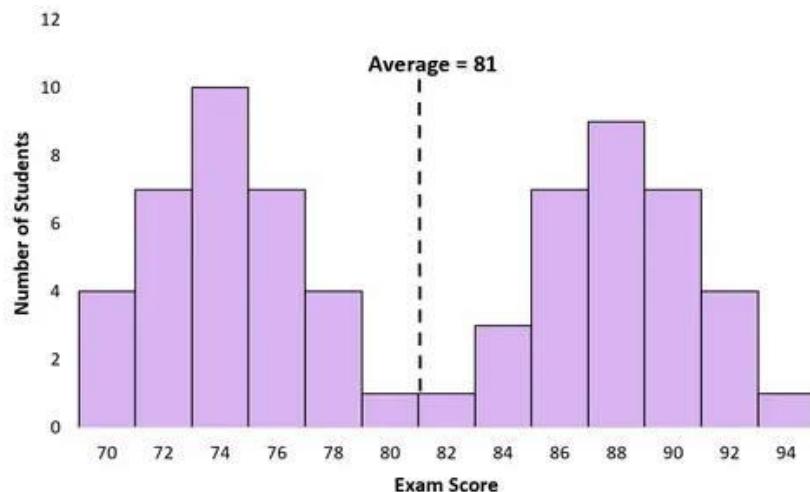
АНЕКДОТ №244368

Согласно статистике, у среднего человека одна грудь и одно [REDACTED]



Среднее арифметическое. Минусы

1. Может трактоваться неоднозначно
2. Плохой показатель типичности



3

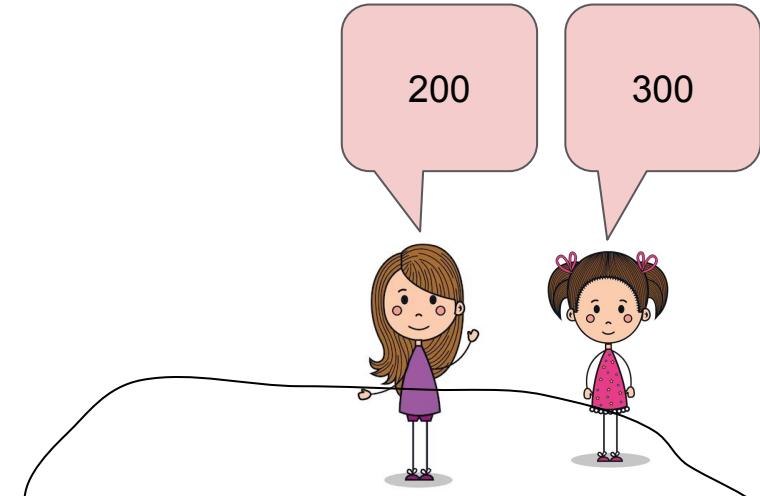
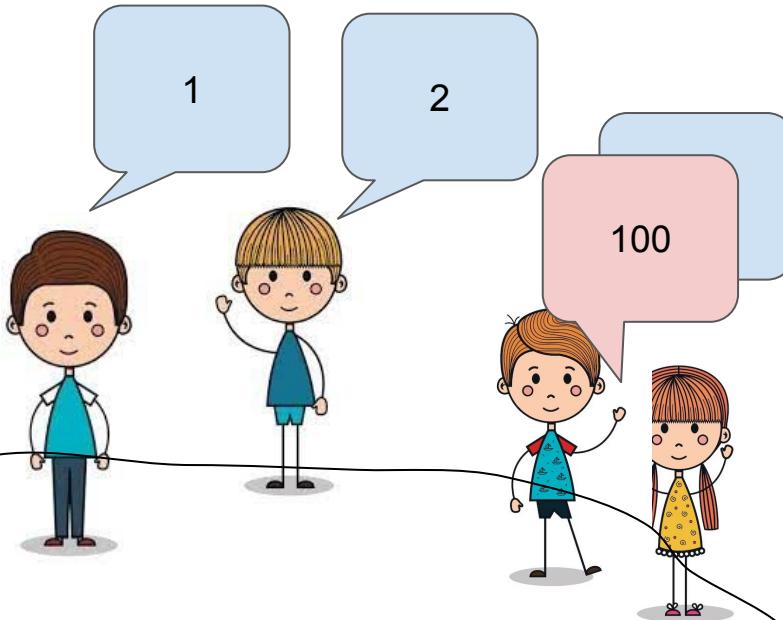


Интуиция

Группа граждан страны А эмигрировала в страну Б.

Мог ли в этих двух странах одновременно вырасти средний IQ?

среднее арифметическое





Феномен Уилла Роджерса

“С 1950 года **больные X** стали жить, в среднем, дольше”



Феномен Уилла Роджерса

“С 1950 года **больные X** стали жить, в среднем, дольше”





Феномен Уилла Роджерса

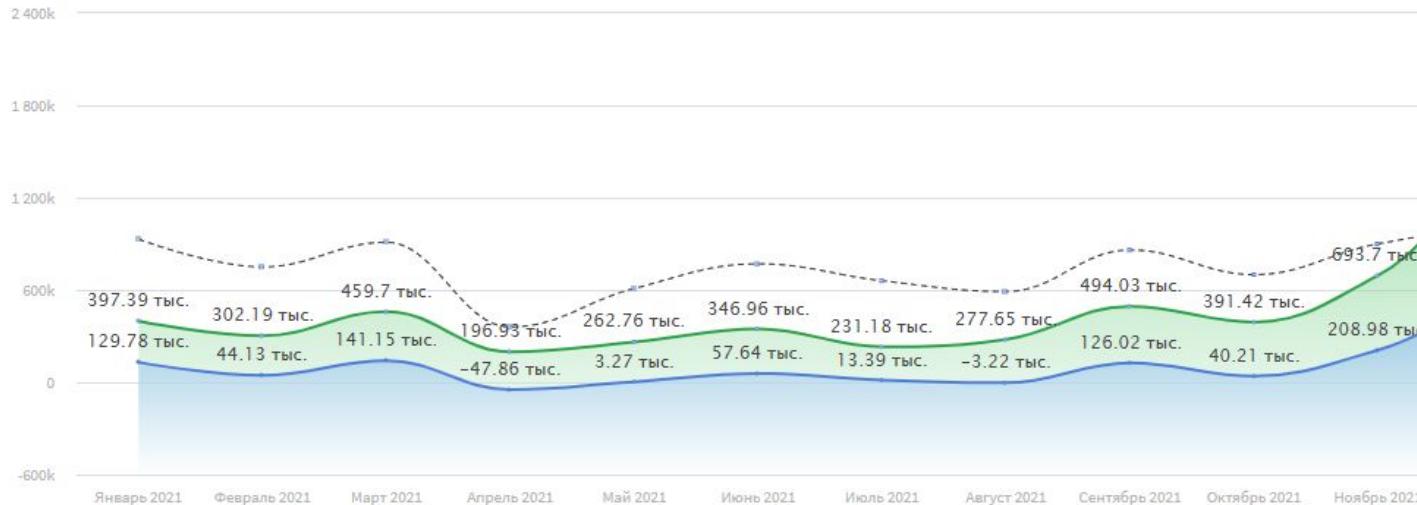
“С 1950 года **больные Х стали жить, в среднем, **дольше**”**





Феномен Уилла Роджерса

По месяцам ▾





Среднее арифметическое. Минусы

1. Может трактоваться неоднозначно
2. Плохой показатель типичности
3. В динамике может вводить в заблуждение

4



Среднее арифметическое. Минусы

1. Может трактоваться неоднозначно
2. Плохой показатель типичности
3. В динамике может вводить в заблуждение
4. Плохо интерпретируем

5

Пример №1

- Журнал “Forbes” хочет построить рейтинг самых высокооплачиваемых профессий.
- Редакторы рассмотрели несколько профессий и узнали зарплаты 5 работников каждой из них.
- Как построить рейтинг? Ваше мнение...

Пример №1

12 8 10 10 10



№1

1 2 3 4 40

№2



Пример №1

- Первое, что приходит в голову: среднее арифметическое.

Пример №1

Посчитайте среднее арифметическое

12 8 10 10 10



№1

1 2 3 4 40

№2



Пример №1

“У нас средняя зарплата 10 тыс. тугриков!”

12 8 10 10 10



№1

1 2 3 4 40



№2

Пример №1

Средняя зарплата одинакова. Но множества совсем разные.

12 8 10 10 10



№1

1 2 3 4 40



№2

Еще приколы среднего арифметического



Еще приколы среднего арифметического



Еще





Среднее арифметическое. Минусы

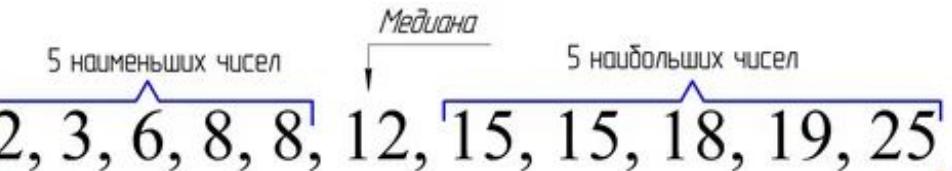
1. Может трактоваться неоднозначно
2. Плохой показатель типичности
3. В динамике может вводить в заблуждение
4. Плохо интерпретируем

Медиана

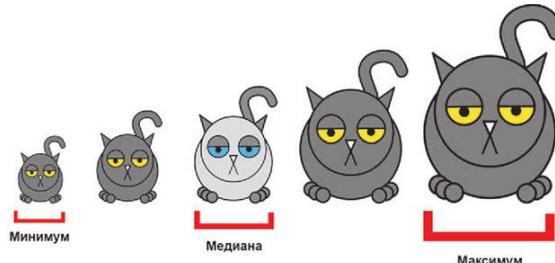
Медиана конечного множества чисел это – ...

Число, которое находится в середине набора чисел, если его упорядочить по возрастанию.

α



Д



С

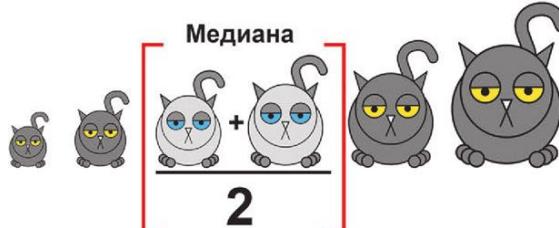
Для множества действительных чисел $S = \{a_1, \dots, a_n\}$ такое число $m \in S$, что $\|\{a_i \in S \mid a_i < m\}\| - \|\{a_i \in S \mid a_i > m\}\| \leq 1$

В

Медиана конечного множества чисел это – ...

Число, которое находится в середине набора чисел, если его упорядочить по возрастанию.

а



д

Для множества действительных чисел $S = \{a_1, \dots, a_n\}$ такое число $m \in S$, что $\|\{a_i \in S \mid a_i < m\}\| - \|\{a_i \in S \mid a_i > m\}\| \leq 1$

с

в

Укажите на медиану

3



1000

Медиана

8 10 10 10 12



№1

1 2 3 4 40

№2



Медиана

50% слева, 50% справа

8 10 10 10 12

10 10



№1

1 2 3 4 40

40



№2

Медиана

Если четное количество

8 10 10 11 12 12



№1

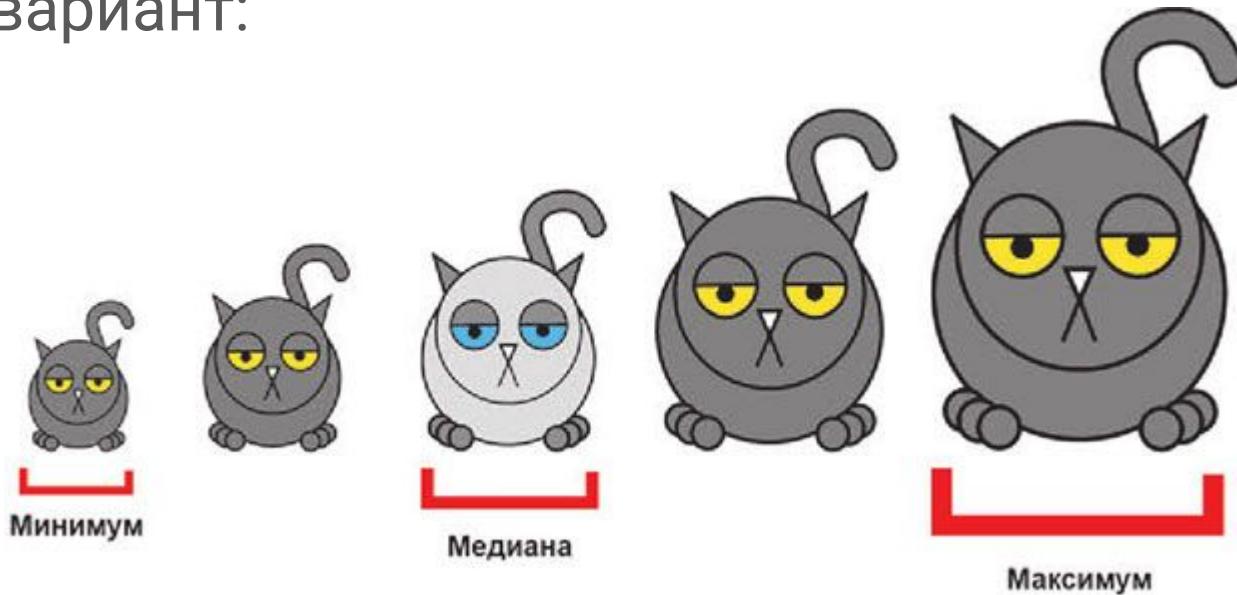
1 2 3 4 40



№2

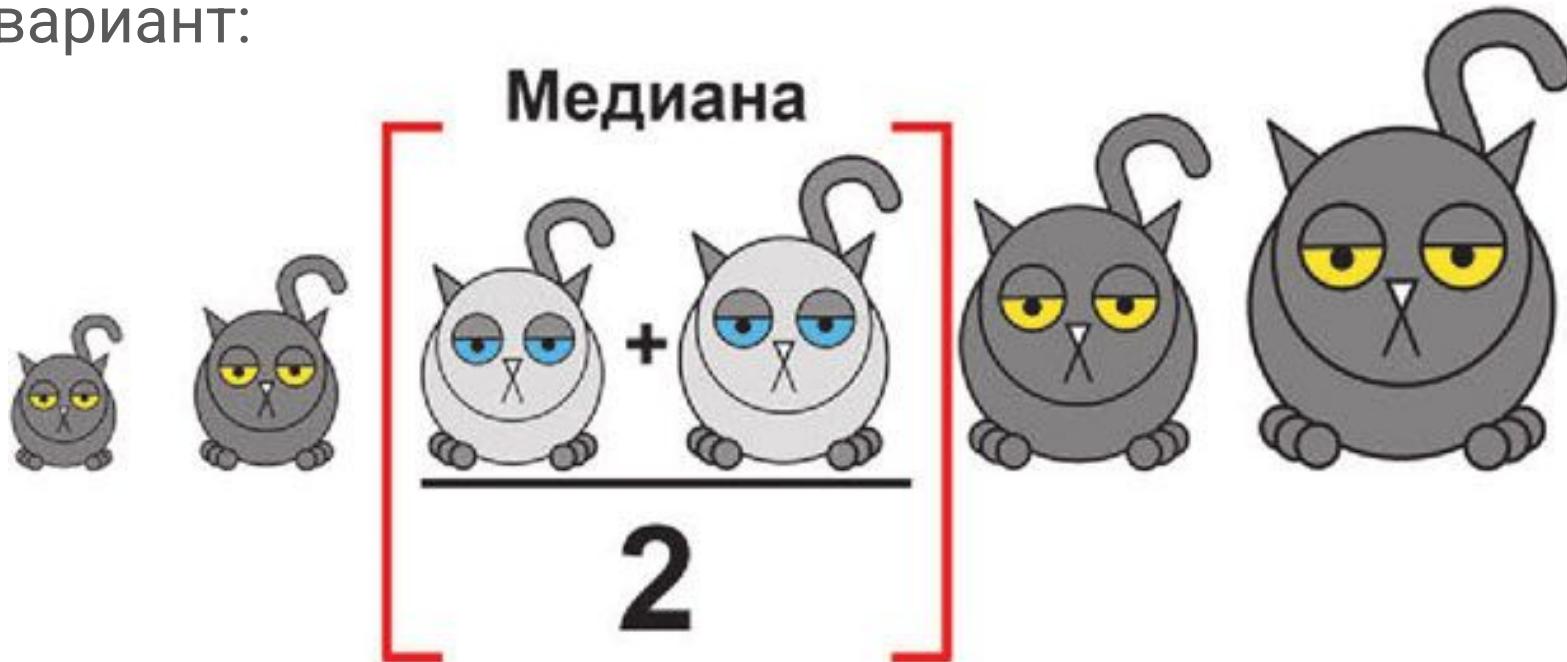
Медиана

Другой вариант:



Медиана

Другой вариант:



Медиана

Свойства

8 10 10 11 12 12



10

11 12



№1

3



№2

Медиана

1. Устойчива к выбросам

8 10 10 11 12 12

10



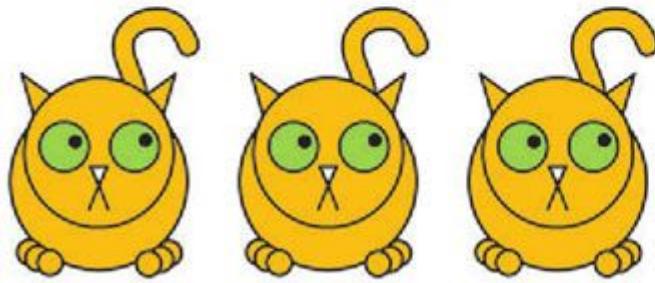
№1

1 2 3 4 40 10000

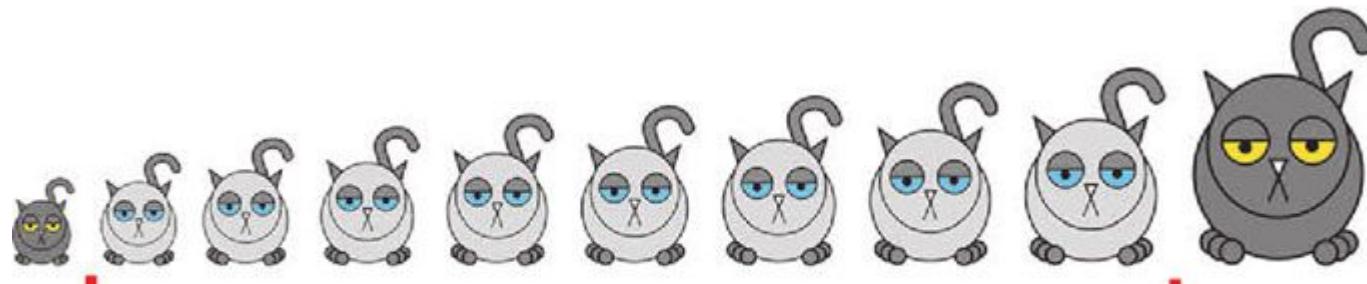
10000



№2



Выброс



Котики для усеченного среднего

Медиана

1. Устойчива к выбросам
2. Указывает на реальный объект

8 10 10 11 12 12

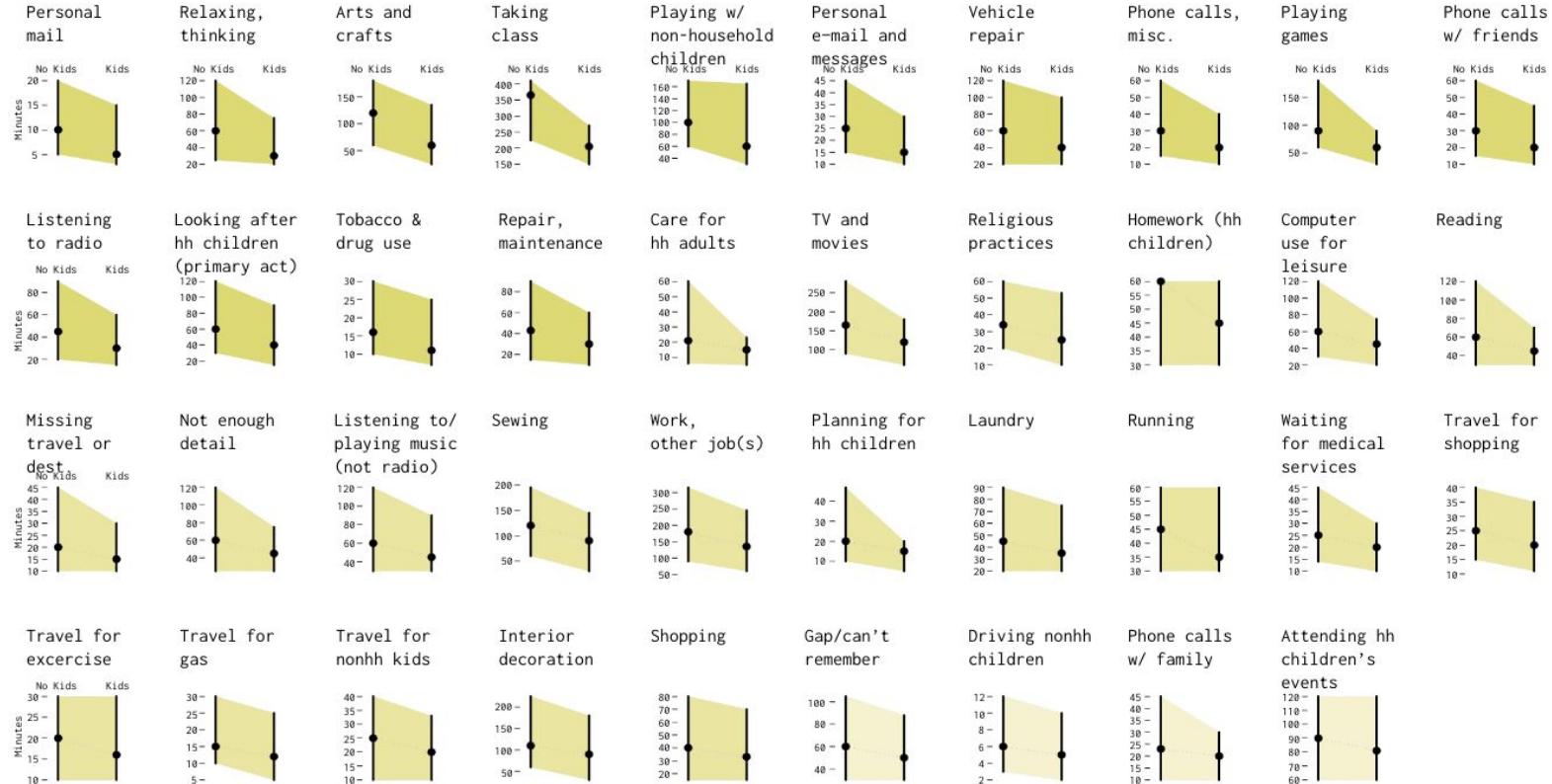


№1

1 2 3 4 40 10000



№2



Шуточки на понимание

11:02 PM 26 11:02 PM 26

For you Following

Find out more

Want more guidance?

David Arakelian @heyiam_david · 1d
Поделюсь радостной новостью и тут. Я уезжаю из Германии. Это были самые худшие полгода в моей жизни. Я переживал очень многое, но просто существование этой страны впервые довело меня до поставленной депрессии. Я не знаю, как здесь живут люди и еще радуются и хотят сюда ехать.

372 161 2,539 374K

Вастринки @vas3k · 2d
СМИ вам этого не расскажут, но в Германии половина населения зарабатывает даже меньше медианной зарплаты по стране!

129 104 2,190 230K

Сензорита фронтенда @dev_margo... · 1d
Никто меня так не бесит, как менторы, которые учат будущих фронтендеров, что ООП и ФП - это классовые и функциональные компоненты в реа

+
Home Search Notifications (1)
Post your reply

← Post

@dimonb@must.beer @dimonb · 2d
Replying to @vas3k
Это смотря четное или нечетное количество людей в Германии живет

1 110 10.5K

Show replies

юрий молодцов ❤️ @yury_mol · 2d
Replying to @vas3k
Вот в советском союзе 80% рабочих зарабатывали больше медианы!

10 1 217 23K

бока кита 🐋 @blessobsession · 2d
Replying to @vas3k
после этой новости моя жизнь развернулась на 360 градусов

1 253 12.5K

Ivan @IvanDotMe · 2d
Replying to @vas3k
Неудивительно, ведь в среднем у жителей Германии меньше 2 рук.

1 49 6,795

Adskiy Sotona @Adskiy_Sotona · 2d
Replying to @vas3k
Это ужасно! Как можно такое скрывать?! Куда смотрит научная общественность?

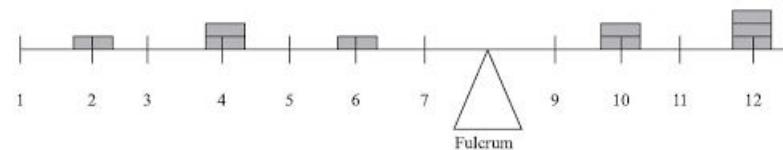
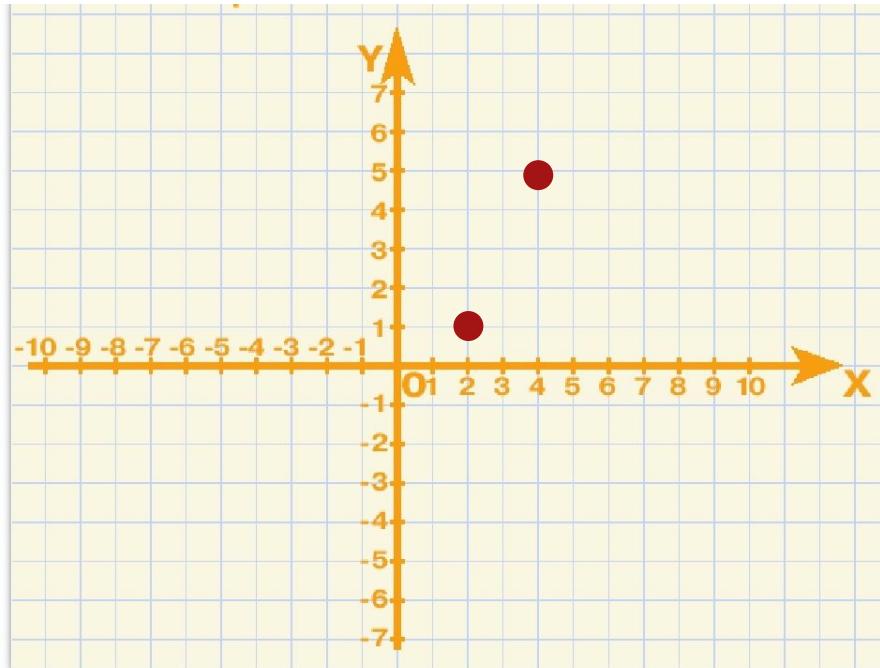
1 15 7,208

Недостатки

Видимо, одного признака мало ...

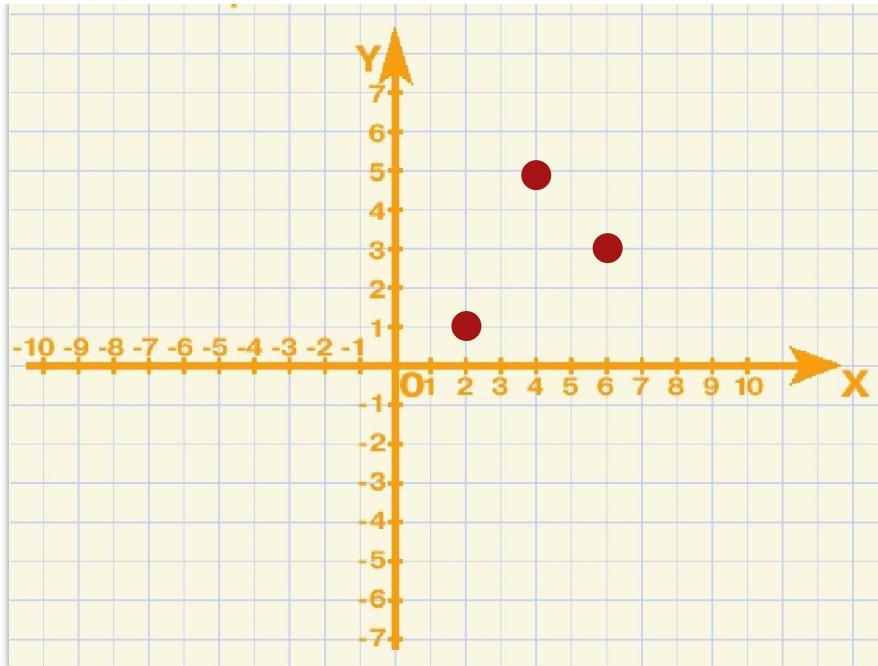
Задача

Среднее арифметическое легко обобщается на \mathbb{R}_2 :



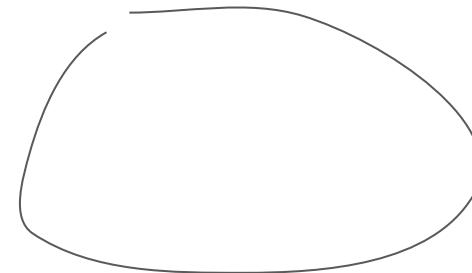
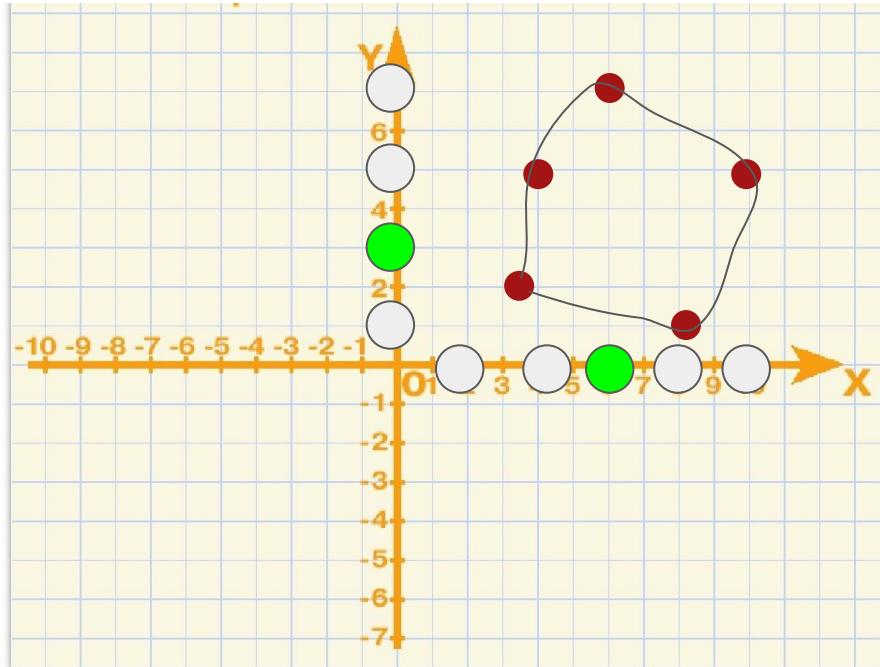
Задача

Среднее арифметическое легко обобщается на \mathbb{R}_2 :



Задача

Как обобщить понятие **медианы** на плоскость (\mathbb{R}_2)?



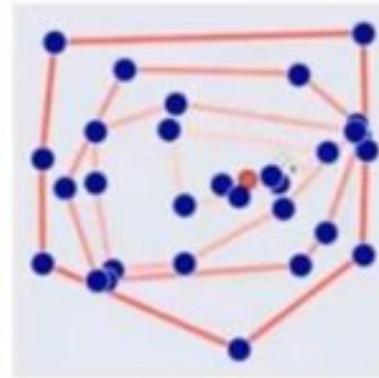
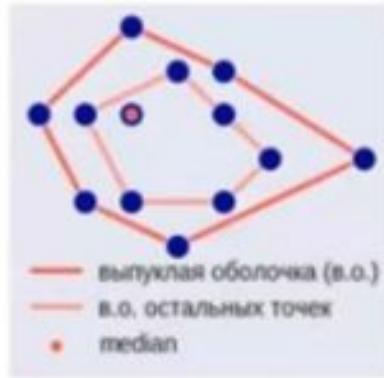
**другие
характеристики**

Задача

Как обобщить понятие медианы на плоскость (\mathbb{R}_2)?

Какие важные свойства будем сохранять?

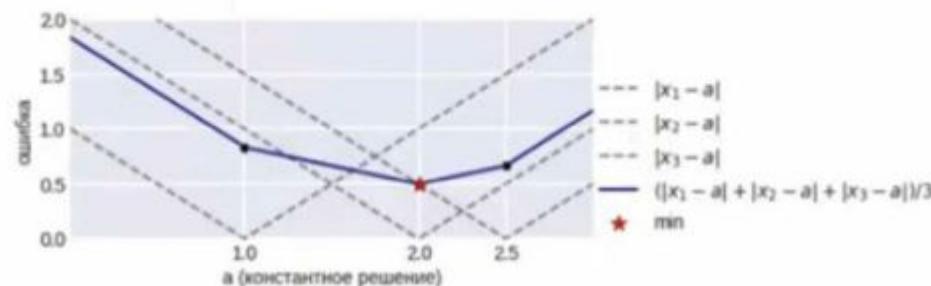
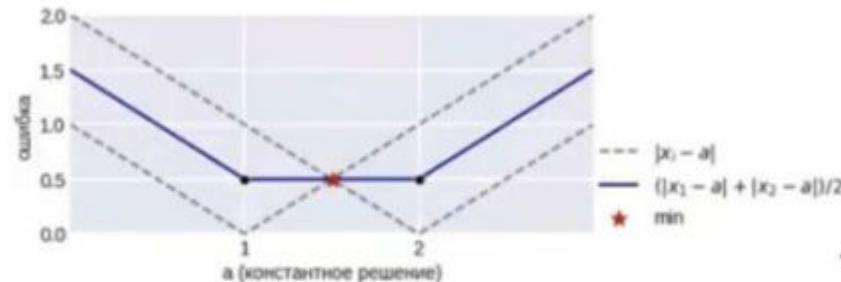
Вариант 1. Обобщить способ построения



**Выход: сделать аналогичный процесс построения,
как в одномерном случае
удаление крайних элементов!**

Вариант 2. Обобщить задачу

Среднее как решение оптимизационной задачи

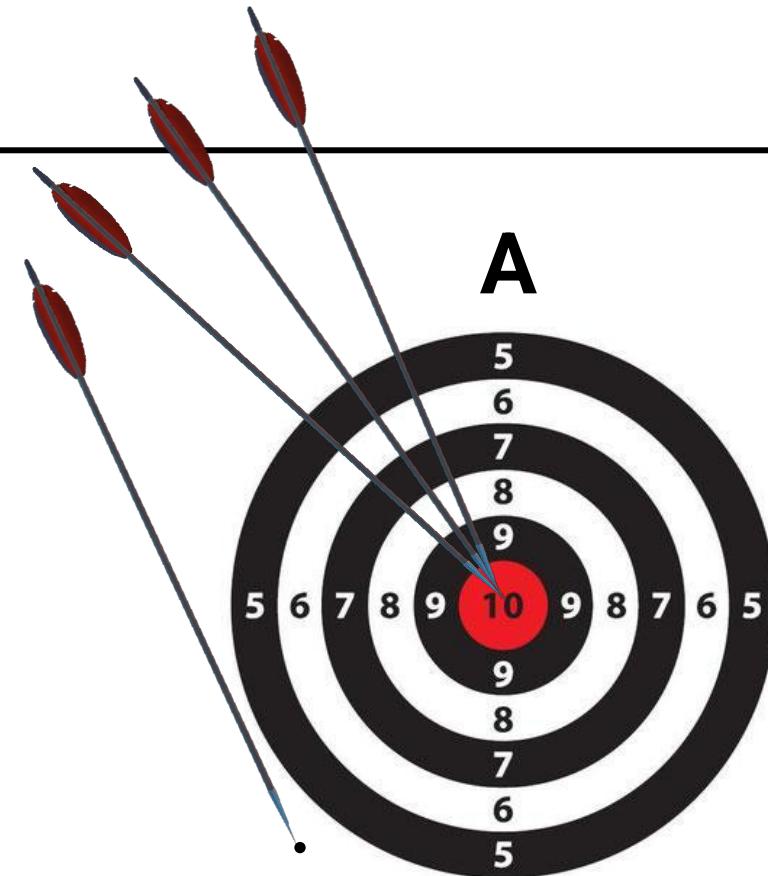


Среднее значение
При наименьшем
ошибке для
Константное
то Минимум
то бы значение
запись

Начните
Быстро

Мода

Одно или несколько значений во множестве, которое встречается наиболее часто.



Типичный россиянин

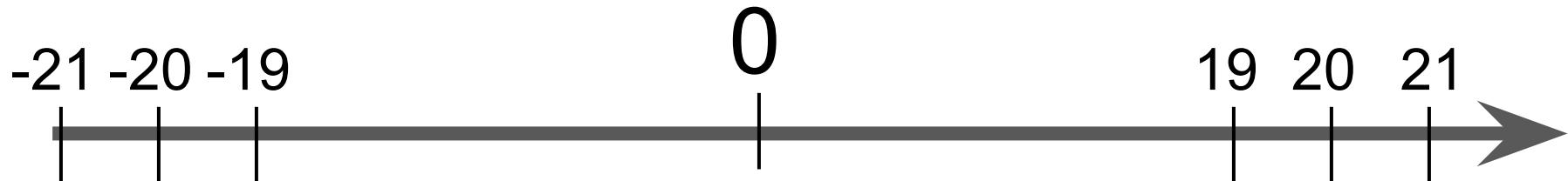
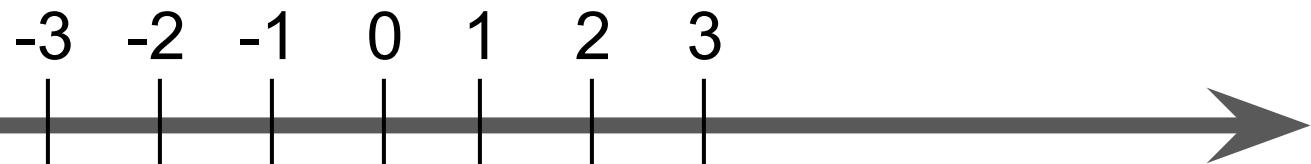
- Женщина
- 38 лет
- Android
- Слушает О. Газманова
- Оптимист

<https://journal.tinkoff.ru/citizen/>



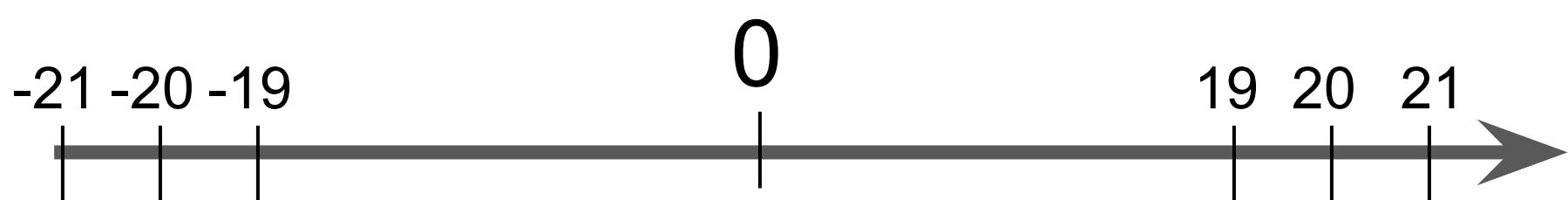
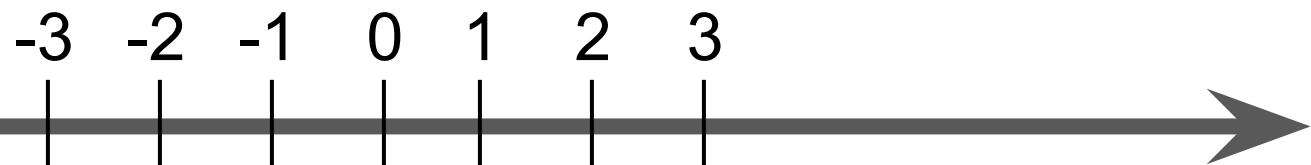
Пример №2

Вы выбираете страну для отпуска. Вот температура за 7 дней в двух странах-фаворитах:



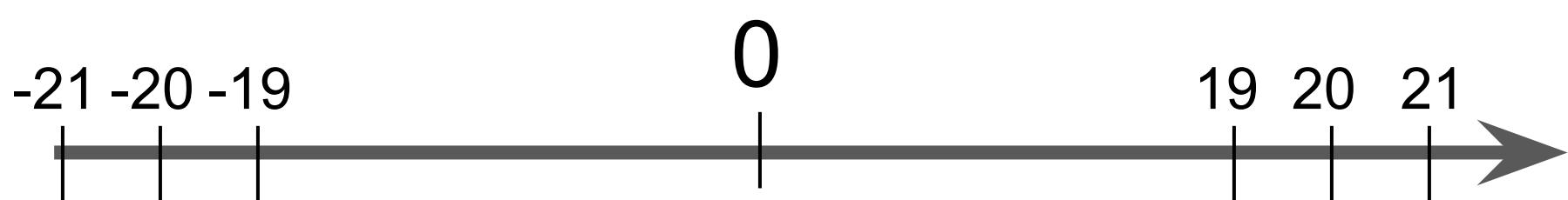
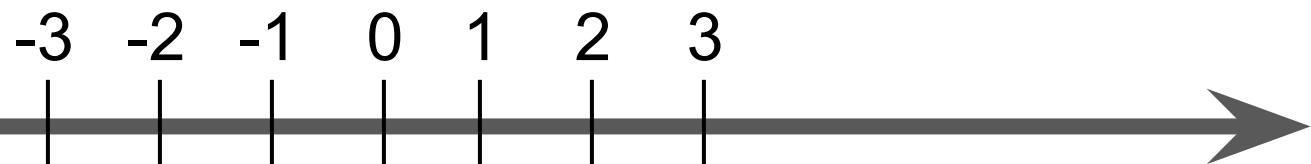
Пример №2

Каково среднее арифметическое?



Пример №2

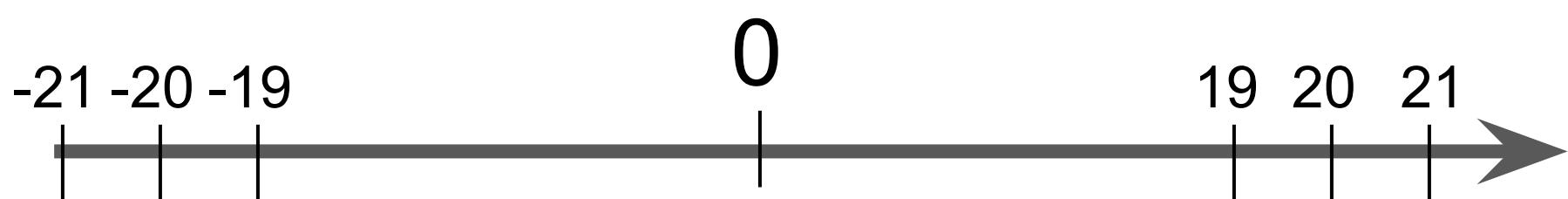
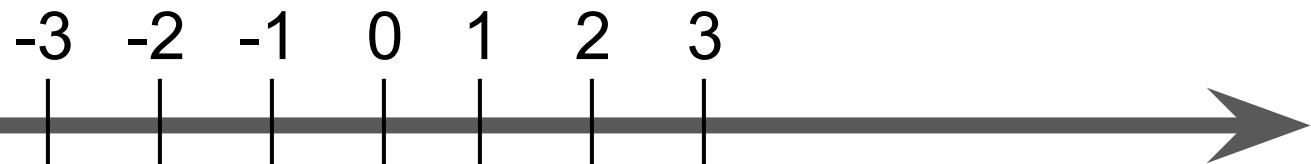
Каково среднее арифметическое? **0**



Пример №2

Каково среднее арифметическое? **0**

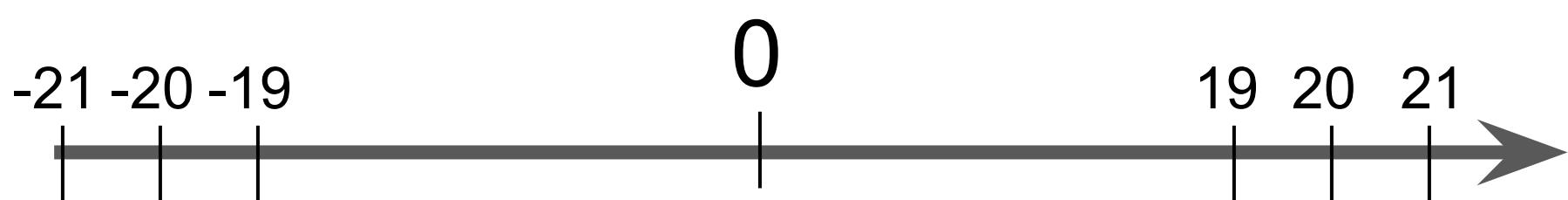
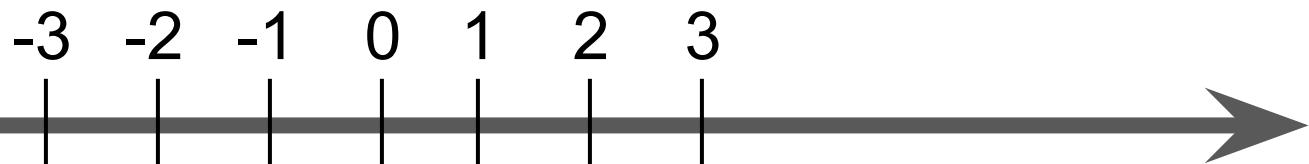
Какова медиана?



Пример №2

Каково среднее арифметическое? **0**

Какова медиана? **0**

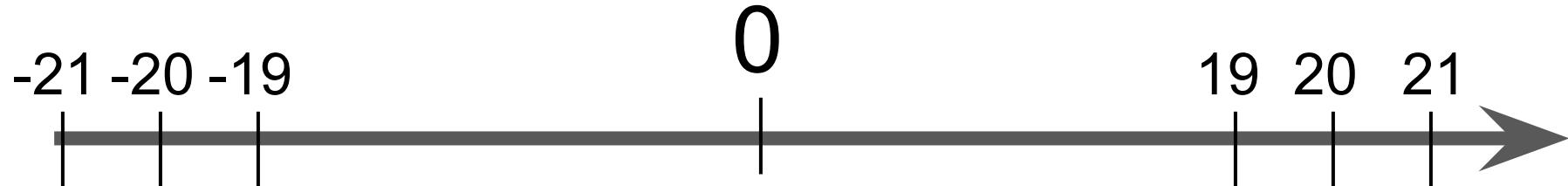
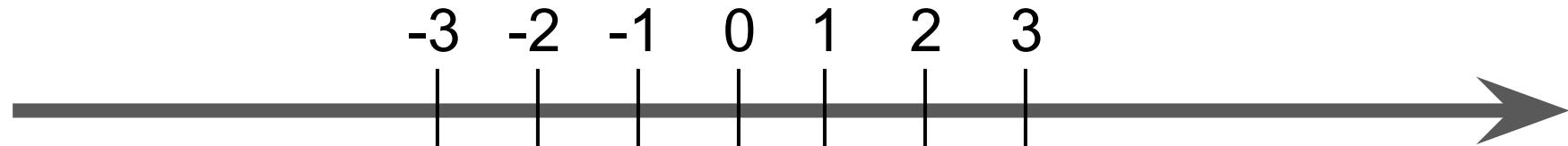


Пример №2

Каково среднее арифметическое? **0**

Какова медиана? **0**

Какую страну выберите?

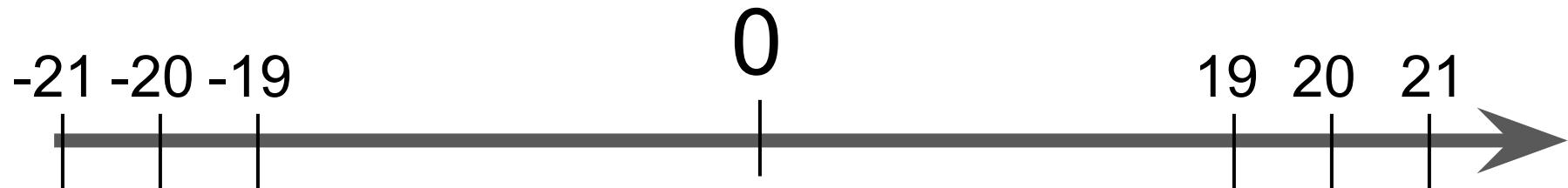
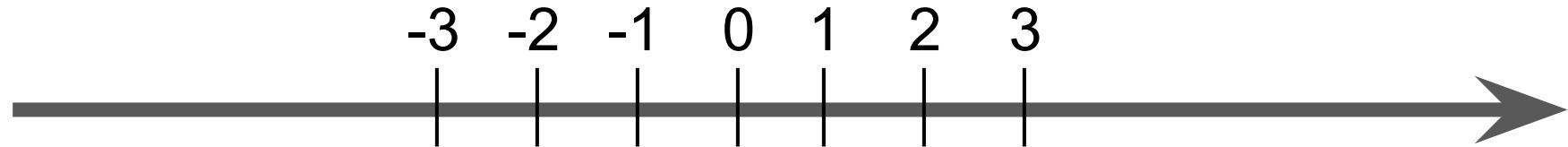


Пример №2

Каково среднее арифметическое? **0**

Какова медиана? **0**

Какую страну выберите? **Какую бы тут характеристику подобрать?**



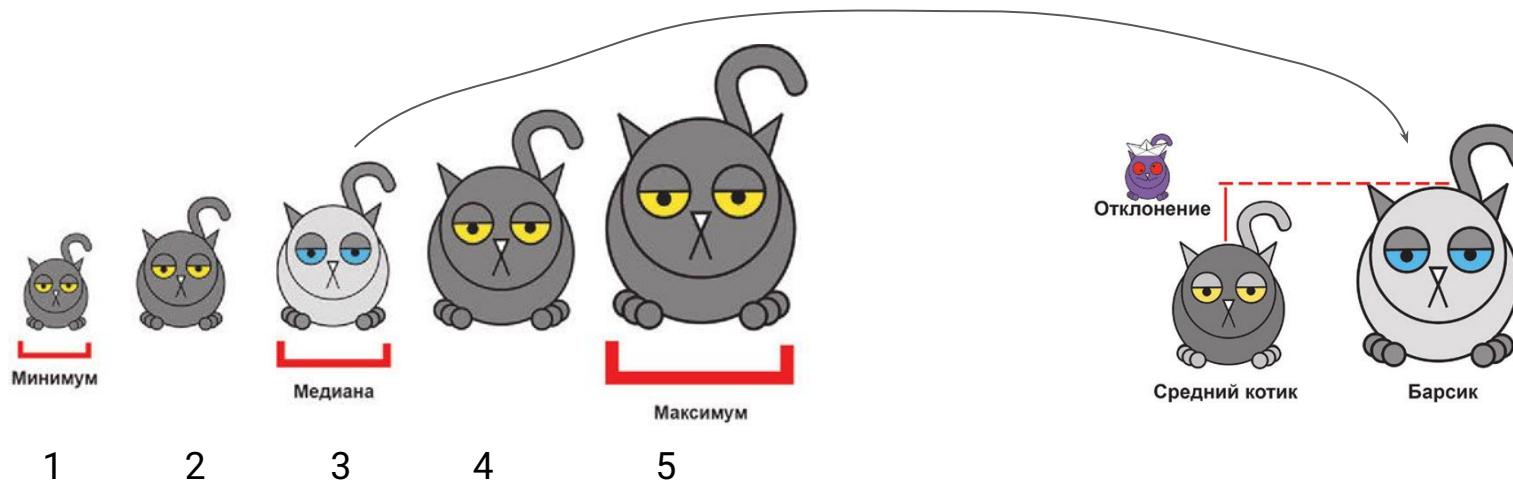
Характеристики

1. Среднее арифметическое
2. Медиана
3. Размах



Характеристики

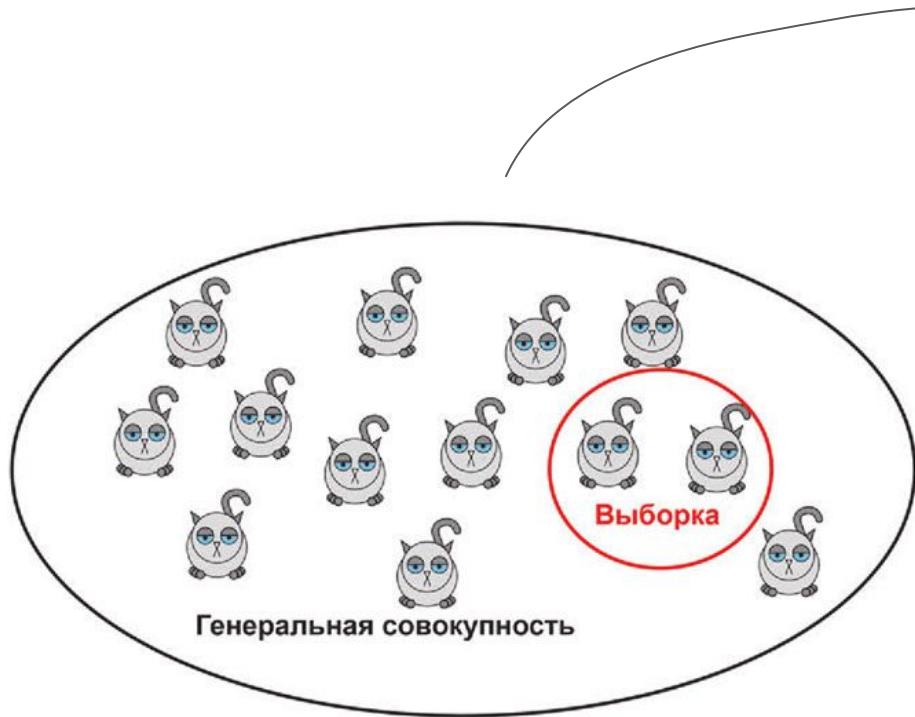
1. Среднее арифметическое
2. Медиана
3. Размах
4. Отклонения. Среднее отклонение считать плохо



Характеристики

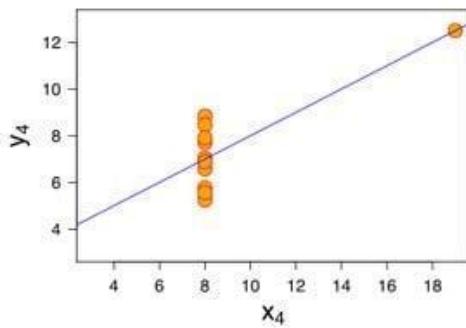
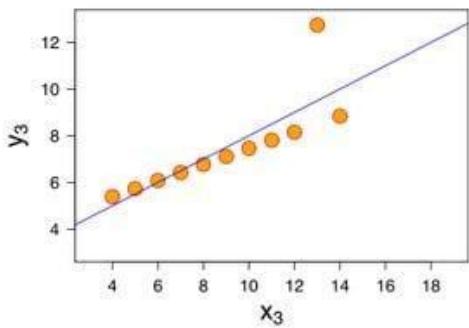
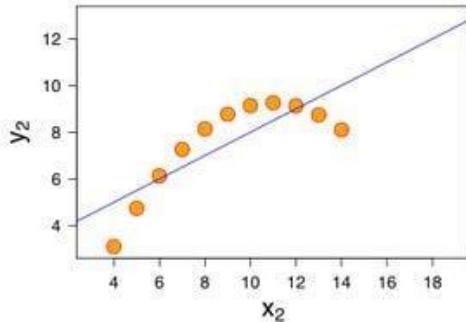
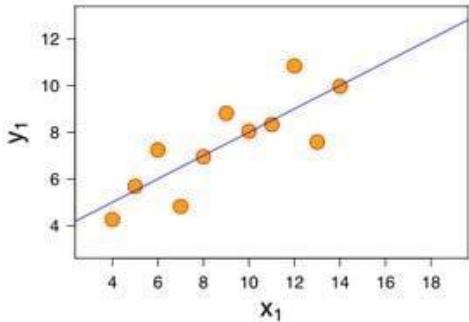
1. Среднее арифметическое
2. Медиана
3. Размах
4. Отклонения:
 - a. Дисперсия генеральной совокупности
 - b. Среднеквадратическое отклонение
 - c. Среднее модулей





Среднеквадратическое
отклонение

Стандартное
отклонение



Зачем нужна визуализация?

У представленных на картинке четырех наборов – одинаковые средние арифметические, медианы, дисперсии и корреляции

Plotly



Plotly

- Интерактивность
- Хорошая документация
 - <https://plotly.com/>
- Просто красиво



Plotly

- Front-End на **JS**
- Back-End на **Python**
- Back-End на **R**



Seaborn



Matplotlib