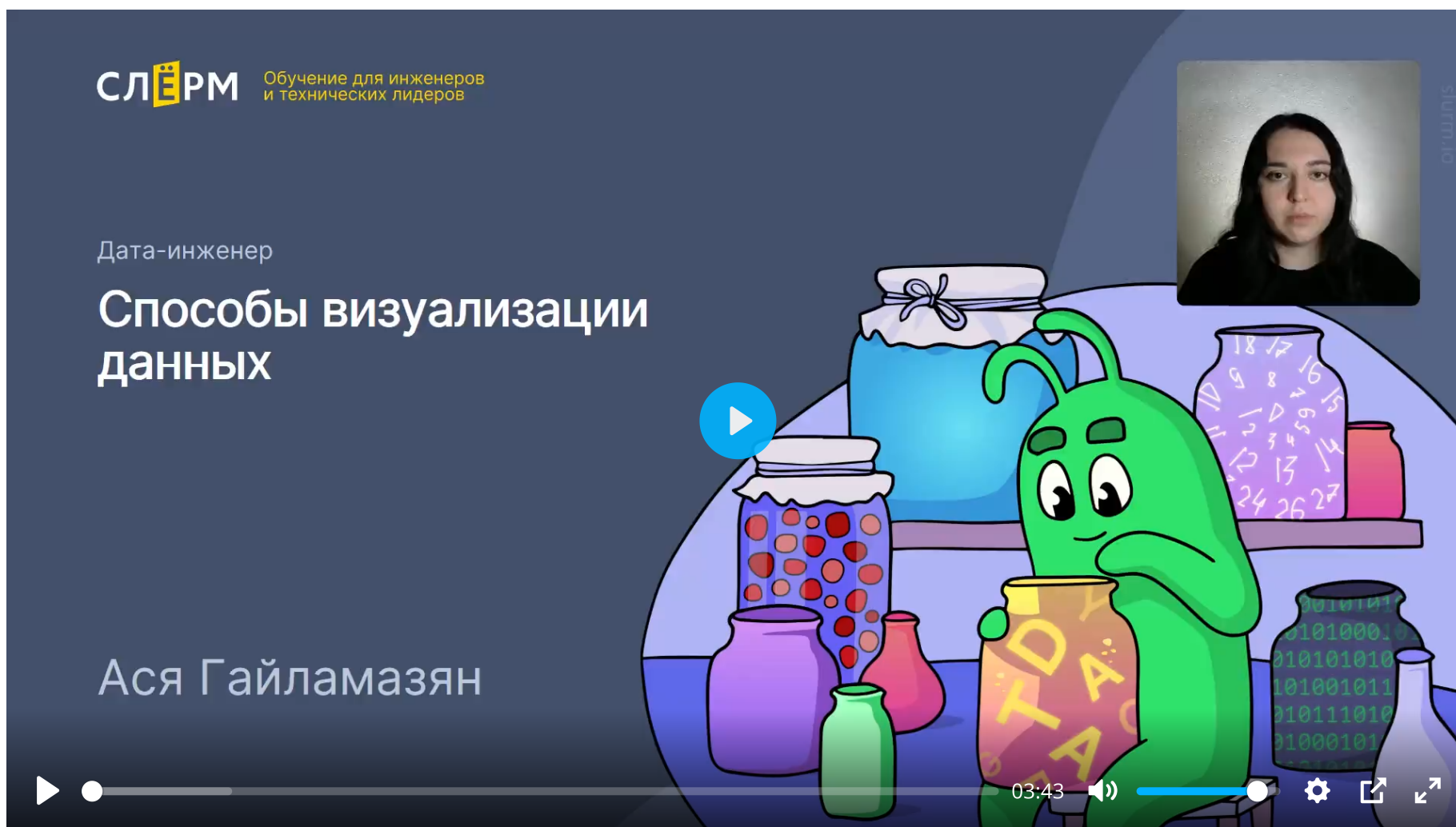




[Презентация к уроку 12.3.2](#)

Текстовая расшифровка видео:

СПОСОБЫ ВИЗУАЛИЗАЦИИ ДАННЫХ

План:

- Что из себя представляет визуализация;
- Диаграмма размаха («ящик с усами»), гистограмма;
- Тепловая карта, диаграмма рассеяния.

ЧТО ИЗ СЕБЯ ПРЕДСТАВЛЯЕТ ВИЗУАЛИЗАЦИЯ

Постановку на визуализацию данных/задач делают люди со стороны бизнеса или аналитики.

Важно помнить:

- **В научной среде** наиболее **главным** считается **самый подробный график**. Побочными являются менее подробные.
- **В бизнесе** – наоборот. **Главным** считается **график с меньшей/лаконичной информацией**, где при переходе в другие места можно увидеть более подробные разбивки.

Одни и те же данные могут представляться разными способами для разных целей. Так, например, **столбчатая диаграмма** хорошо подходит для изображения дискретных данных, а, например, **непрерывные графики** нужны для изображения данных с большей частотой разбивок (по неделям, по дням).

<https://datavizproject.com/>

<https://datavizcatalogue.com/RU/>

Рассмотрим менее распространенные диаграммы.

Диаграмма размаха («ящик с усами»), гистограмма

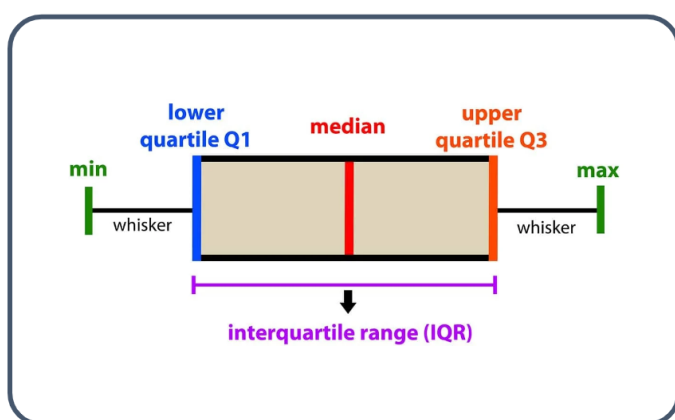
Диаграмма размаха («ящик с усами») позволяет наглядно отображать:

- медиану или среднее значение;
- нижний и верхний квантили;
- минимальные и максимальные значения выборки;
- выбросы.

Их можно располагать вертикально, горизонтально, а также ставить несколько диаграмм рядом. Расстояния между частями этого ящика показывают степень разброса (дисперсия в статистике) и асимметрию данных, также позволяют выявить выбросы.

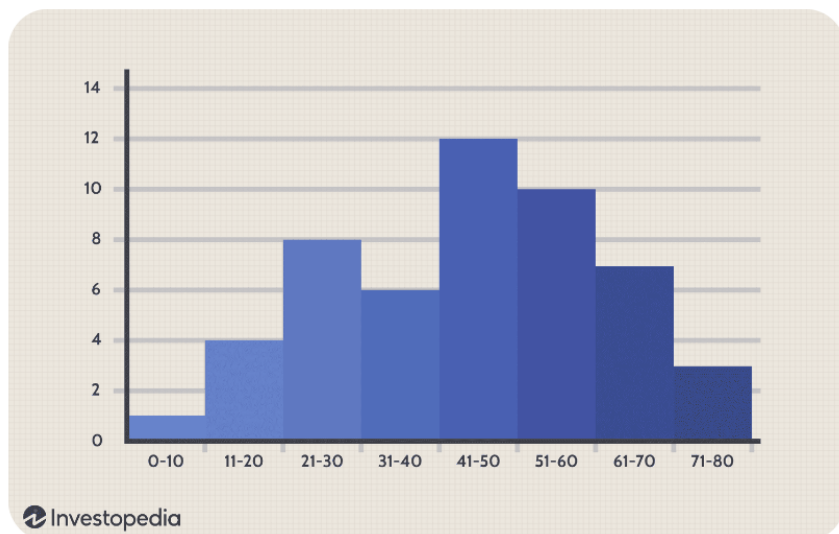
Медиана – срединное значение набора, то есть, если упорядочить все значения в наборе по возрастанию, то одна половина значений будет не больше, чем медиана, другая же – не меньше, чем медиана.

Квантиль – это такое же значение. Однако для него 25% значений не превышают его, а 75% – не менее, чем нижний квантиль. Также и для верхнего квантиля: 75% – не больше, 25% – не меньше.



Гистограмма отличается от столбчатой диаграммы следующим:

- Столбчатая диаграмма – это изменения для одной переменной;
- Гистограмма показывает определение данных. Это тоже статистические данные.



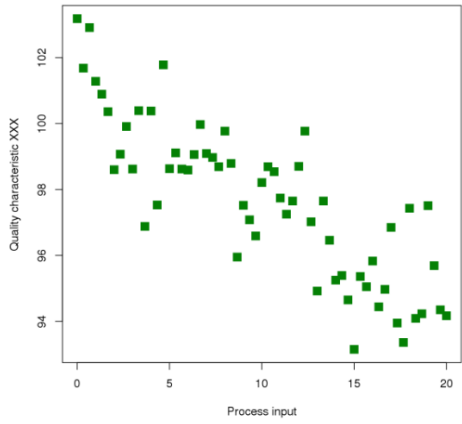
Тепловая карта, диаграмма рассеяния

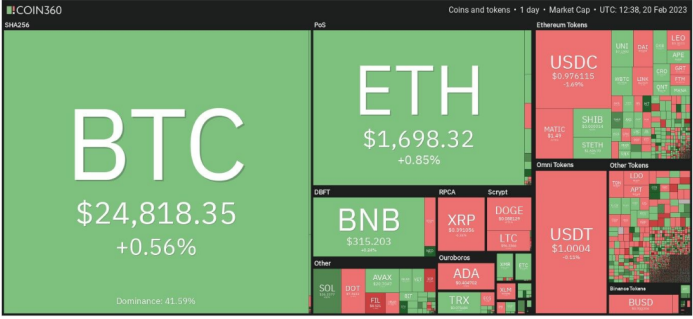
Тепловая карта – это графическое отображение данных, на которых различные данные отображаются с помощью разницы в цвете.

Matplotlib, Seaborn, Plotly и некоторые другие библиотеки Python позволяют рисовать тепловые карты. Они используются для того, чтобы, например, отображать активность на сайте (яркие места – отображение того, куда пользователи чаще всего нажимали), а также, например, для курсов валют (зеленый цвет в значении «+», красный в значении «падение», но бывают и различающиеся по цветам графики).

Точечная диаграмма (Scatter Plot) – это способ представления данных, при котором каждый объект представляется точкой по оси «X» и по оси «Y». Такой график очень полезен для отображения распределения данных (можно найти кучкования данных, их группировки) и сегментирования.

Scatterplot for quality characteristic XXX





Как вам урок?

☹️

😐

🔥

Изучил, далее >