

KARPOV.COURSES >>>

КОНСПЕКТ



## > Конспект > 1 урок > ВИЗУАЛИЗАЦИЯ

### > Оглавление урока по Визуализации

1. Зачем нужны графики?
2. Line chart
3. Bar chart
4. Pie chart
5. Area chart
6. Scatter plot
7. KPI и Фактоиды
8. Таблица
9. Подключения и базы данных в Tableau
10. Интерфейс Tableau
11. Основные понятия
12. Материалы лекции

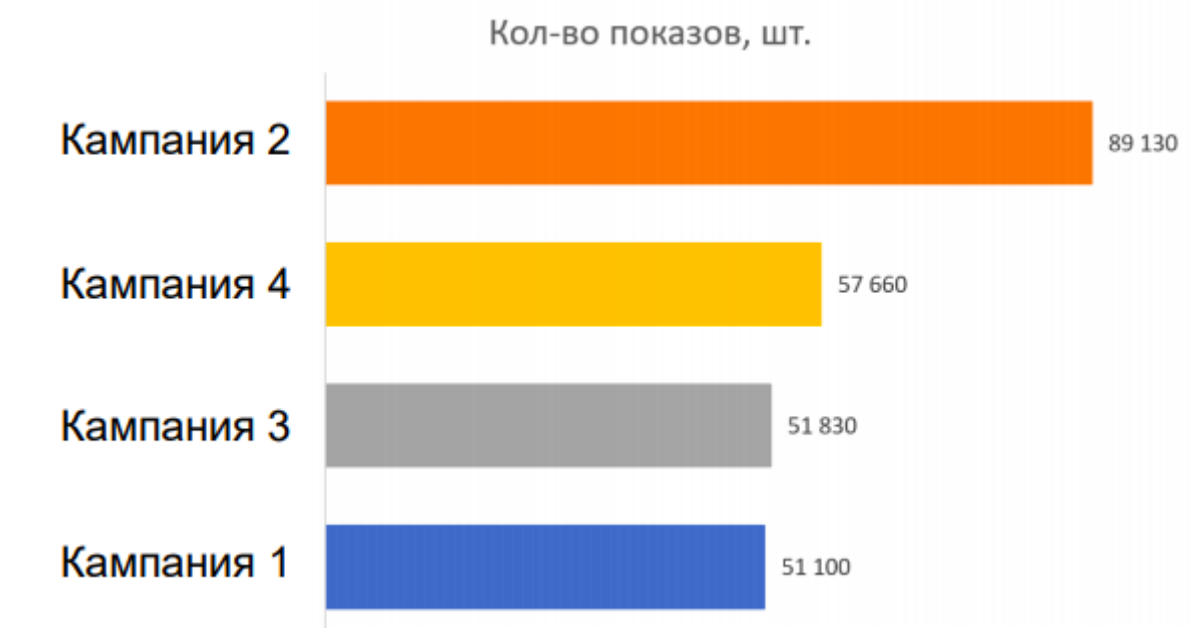
## > Зачем нужны графики?

### 1. Визуализация хорошо отвечает на разные аналитические вопросы

Давайте посмотрим на примере. Предположим, у нас имеются данные о показах рекламного баннера в четырех разных кампаниях. Какая кампания заняла третье место по просмотрам?

Дата	Кол-во показов рекламного баннера				Итого:
	Кампания 1	Кампания 2	Кампания 3	Кампания 4	
14.02.2024	2700	2500	2498	1050	8748
15.02.2024	2750	2740	2496	3850	11836
16.02.2024	2200	2590	2503	1050	8343
17.02.2024	2200	2500	2500	4010	11210
18.02.2024	2950	2800	2505	2350	10605
19.02.2024	2300	3580	2500	1000	9380
20.02.2024	2350	2920	2300	4400	11970
21.02.2024	2250	4660	2450	1650	11010
22.02.2024	2850	2500	2500	4650	12500
23.02.2024	2050	4600	2490	2750	11890
24.02.2024	2350	4480	2000	1500	10330
25.02.2024	2300	3580	2500	3650	12030
26.02.2024	2900	3280	2300	1200	9680
27.02.2024	2350	4600	2459	2950	12359
28.02.2024	2400	7000	2470	2600	14470
29.02.2024	2200	5440	2700	3600	13940
01.03.2024	2050	7600	2500	2600	14750
02.03.2024	2100	6280	2400	2900	13680
03.03.2024	3000	4780	2480	1050	11310
04.03.2024	2350	5600	2800	2650	13400
05.03.2024	2500	5100	2479	6200	16279
Итого:	51100	89130	51830	57660	249720

А теперь посмотрим на график и попробуем ответить на тот же вопрос:

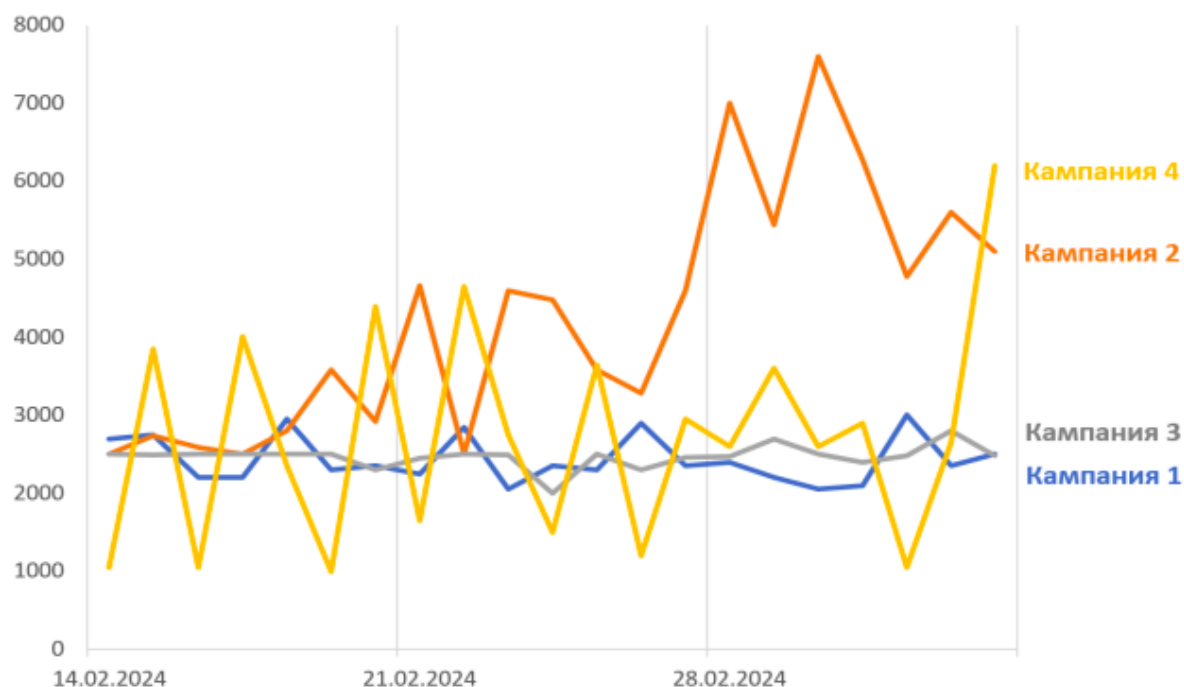


В данном случае ответ найти получается намного быстрее!

Более сложный пример: Какая кампания имеет выраженную положительную динамику за весь период? Какая наиболее волатильна?

Кол-во показов рекламного баннера					
Дата	Кампания 1	Кампания 2	Кампания 3	Кампания 4	Итого:
14.02.2024	2700	2500	2498	1050	8748
15.02.2024	2750	2740	2496	3850	11836
16.02.2024	2200	2590	2503	1050	8343
17.02.2024	2200	2500	2500	4010	11210
18.02.2024	2950	2800	2505	2350	10605
19.02.2024	2300	3580	2500	1000	9380
20.02.2024	2350	2920	2300	4400	11970
21.02.2024	2250	4660	2450	1650	11010
22.02.2024	2850	2500	2500	4650	12500
23.02.2024	2050	4600	2490	2750	11890
24.02.2024	2350	4480	2000	1500	10330
25.02.2024	2300	3580	2500	3650	12030
26.02.2024	2900	3280	2300	1200	9680
27.02.2024	2350	4600	2459	2950	12359
28.02.2024	2400	7000	2470	2600	14470
29.02.2024	2200	5440	2700	3600	13940
01.03.2024	2050	7600	2500	2600	14750
02.03.2024	2100	6280	2400	2900	13680
03.03.2024	3000	4780	2480	1050	11310
04.03.2024	2350	5600	2800	2650	13400
05.03.2024	2500	5100	2479	6200	16279
<b>Итого:</b>	<b>51100</b>	<b>89130</b>	<b>51830</b>	<b>57660</b>	<b>249720</b>

Посмотрев на график, мы сразу же можем понять, что кампания 2 явно имела положительную динамику, а кампания 4 наиболее волатильна и колеблется по просмотрам чаще всего.



Когда мы говорим о визуализации данных, мы понимаем, что различные типы визуализаций хорошо отвечают на какие-то конкретные вопросы. При этом бизнес-пользователи зачастую хотят видеть именно таблицы, т.к. они универсальны (да, они тоже считаются типом визуализации!).

## 2. Визуализация «проявляет» природу данных

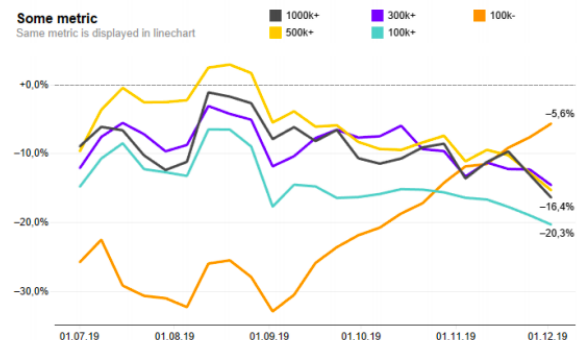
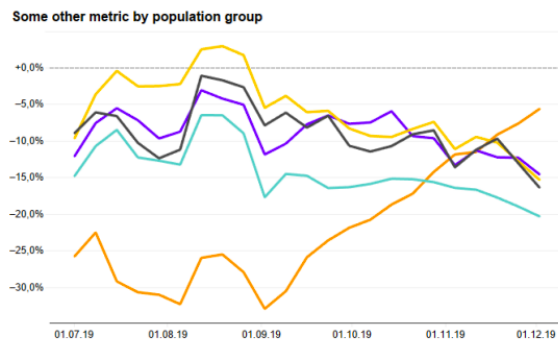
Вторая задача, которая решается с помощью визуализации данных – проявление природы данных. Exploratory data visualization – анализируем данные и пытаемся понять, что в них происходит, перед тем как делать дальнейший анализ.

### > Line chart

Стандартный способ показать изменения данных во времени.

- Можно делать с двумя осями, но лучше с одной.
- Можно поставить маркер на точке, на которой нужно сделать акцент или подписать конец/начало линии.

- Нельзя подписывать все значения точек на графике.

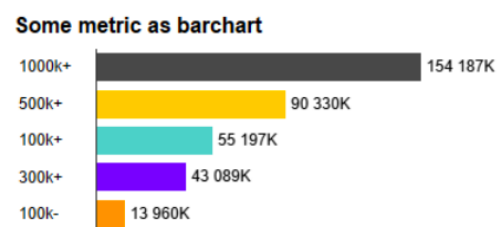
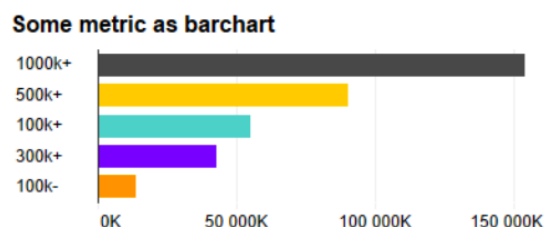


Также не рекомендовано использовать данный тип графика, когда по оси x расположено не время.

## > Bar chart

Хорошо подходит чтобы сравнить долю разных категорий или ранк (топ-N значений).

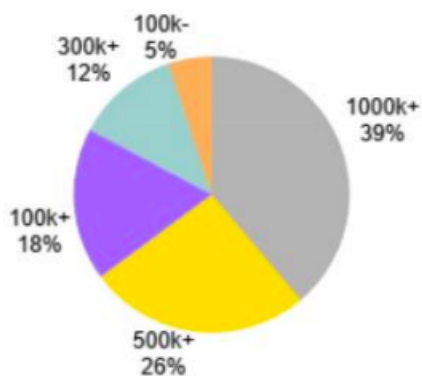
- Лучше использовать горизонтальные бар.
- Вертикальные рекомендуется использовать только, если по оси X — время.



## > Pie chart

Хорошо подходит только для одной задачи: показать соотношение для небольшого количества категорий, образующих одно целое.

**Some title**  
Some description



**Some title**  
Some description

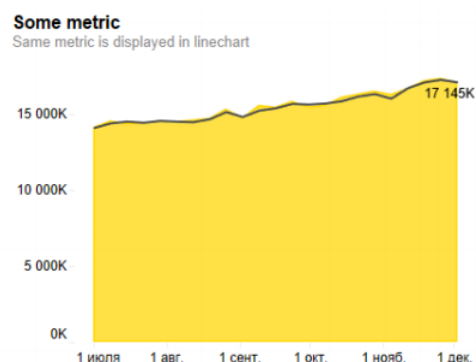
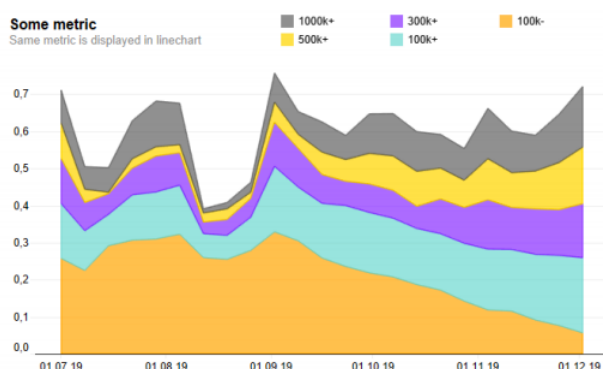


Почему этот тип графика критикуется? Определить точные значения по пай чарту довольно проблематично: считывание угла довольно сложная операция и на глаз воспринимается плохо.

## > Area chart

Такой график используется, когда важно видеть суммарный тренд, а также примерное распределение составляющих компонент.

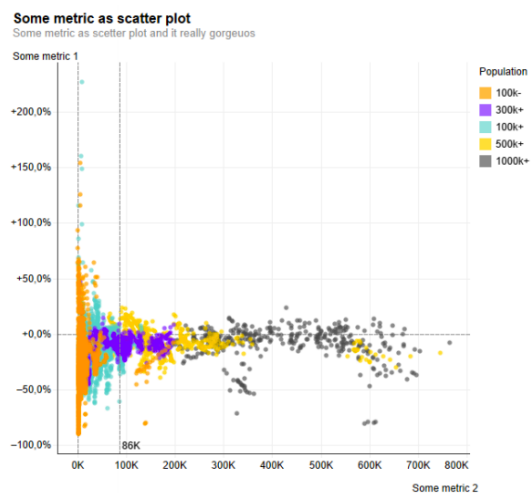
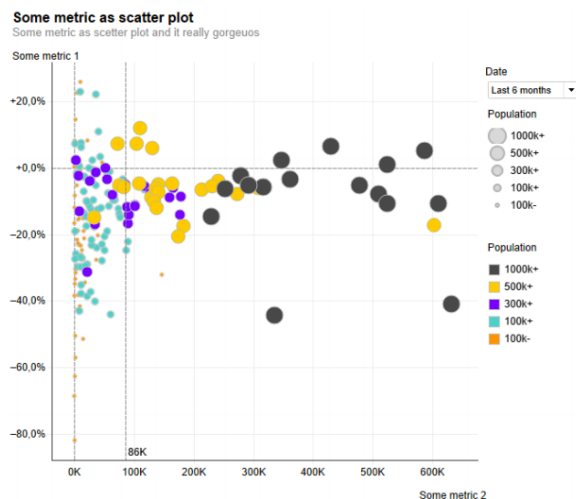
- Внизу располагается наибольшая или наиболее важная компонента.
- При комбинации с line chart, можно показывать плановые и фактические значения.



Частая ошибка – предположение, что по данному графику *"легко отследить динамику каждого компонента"*. Это не так: график хорошо работает для отслеживания общей динамики и понимания примерной разбивки. Например, определить точные изменения динамики желтого сектора практически невозможно.

## > Scatter plot

График для отражения корреляции между двумя величинами. Размерами точки можно показать, например, удельный вес для этой точки.



## > KPI и Фактоиды

KPI и Фактоиды – крупные цифры, которые обозначают текущее значение метрики и тренд за какой-либо период.

- Самое важное значение лучше делать самым крупным.
- Показывать название метрики, которая отображается.
- Можно показывать дополнительную информацию (например, прирост в процентах).

02 Dec ... 08 Dec 19

**Metric Name**

**196K**

▲ +4.5% WoW  
188K +8.44K

**Some KPI with sparkline**

Last 7 days

**17 145K**

▲ +7% Wow



**Some KPI with sparkline**

Last 7 days

**500k+**

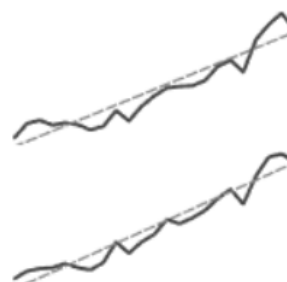
**4 346K**

▲ +7% Wow

**1000k+**

**7 554K**

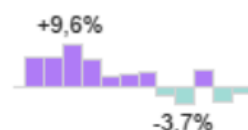
▲ +7% Wow



**Supply Change**

11.11.2019

**-1,4%**



Полезны для отображения текущих значений и трендов, особенно в операционных дашбордах.

## > Таблица

Хорошо позволяет считать точное значение или сравнить несколько разрезов по нескольким метриками.

Population	Some metric 1	Some Metric 2
1000k+	-8,9%	154 187K
500k+	-6,0%	90 330K
300k+	-8,8%	43 089K
100k+	-14,2%	55 197K
100k-	-21,7%	13 960K
Total	-14,0%	356 763K

## > Подключения и базы данных

Итак, одним из первых шагов после открытия Tableau будет подключение к базе данных. Присоединиться можно к:

- Файлу (to a file)
- Серверу (to a server)

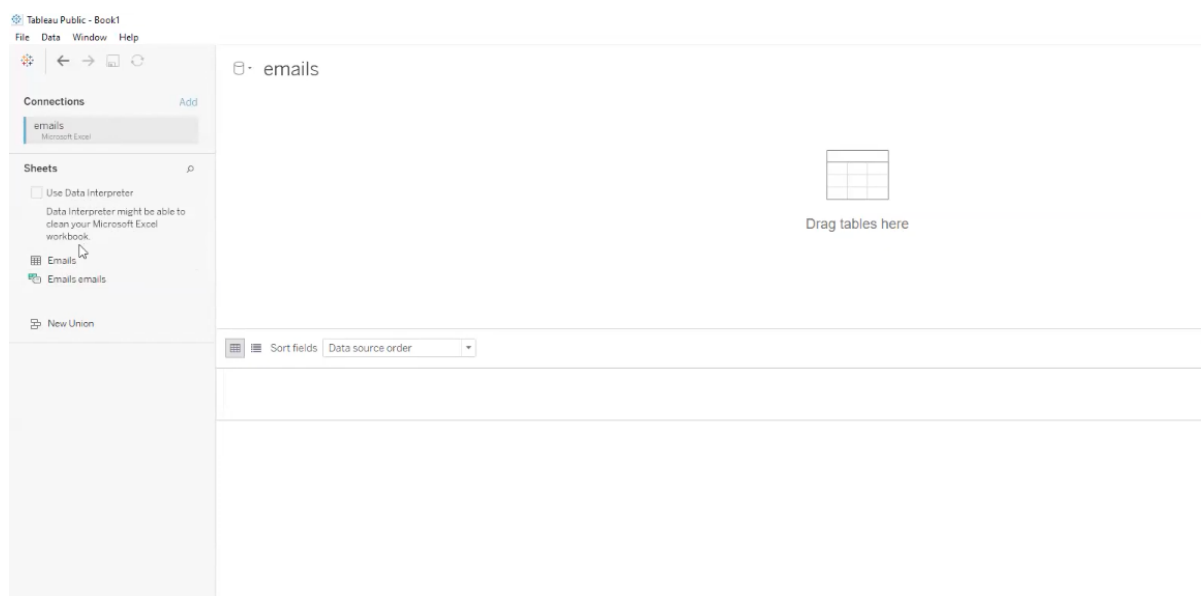
Как правило, обычно подключаются к различным серверам.



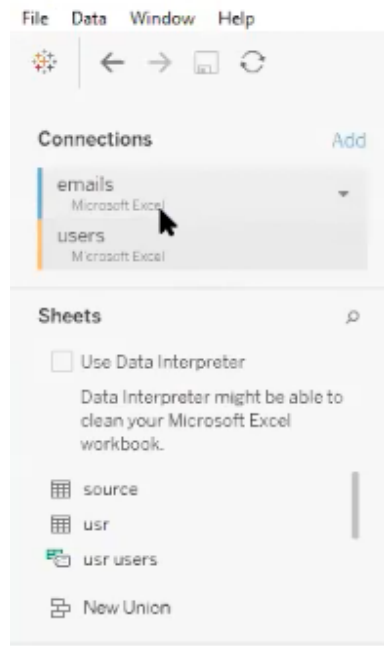
Note: В бесплатной версии Tableau Public не поддерживается подключение к онлайн данным.

## Пример: подключение к таблице

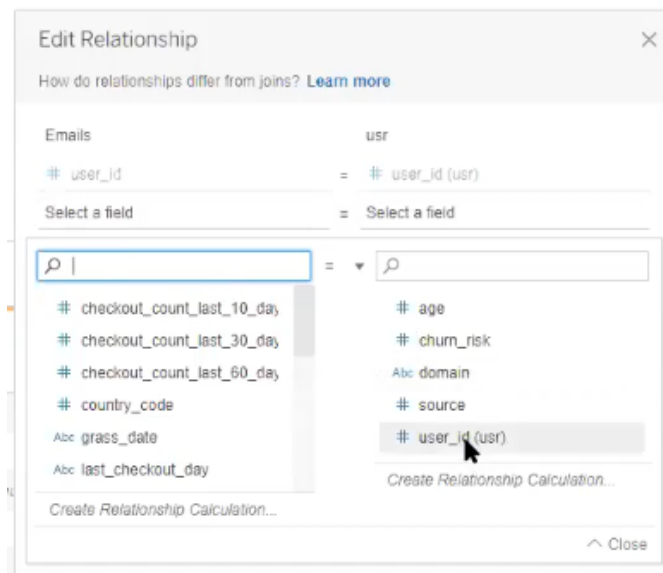
Сначала подключаемся к таблице Microsoft Excel (to a file → microsoft excel → выбираем нужную). Для того, чтобы использовать несколько листов, нужно перетащить нужные листы (sheets) в поле "drag tables here".



Перетащим туда лист Emails – появится первая таблица. И подключимся еще к одному источнику данных, в данном случае это тоже файл xlsx (connections → add → выбираем нужный).

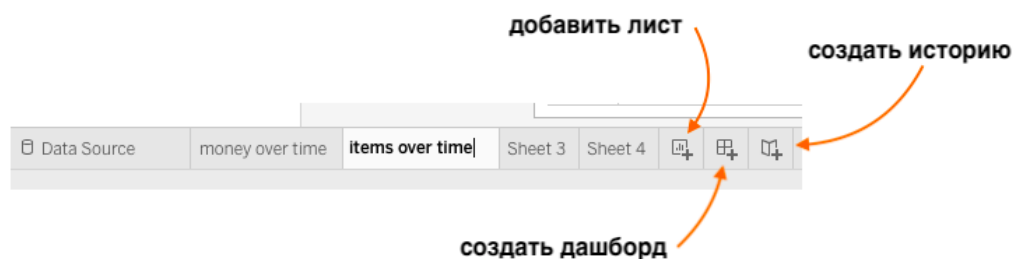


Следующая задача – совместить данные из users и emails. Чтобы сделать это в Tableau, нужно перетащить вторую табличку уже в знакомое нам поле, а затем установить отношения между таблицами (*edit relationship*). Tableau само находит поля с одинаковыми названиями, но также можно выбрать ключи самостоятельно (add field). Это очень напоминает JOIN, но физически отношения таблицы не объединяют.

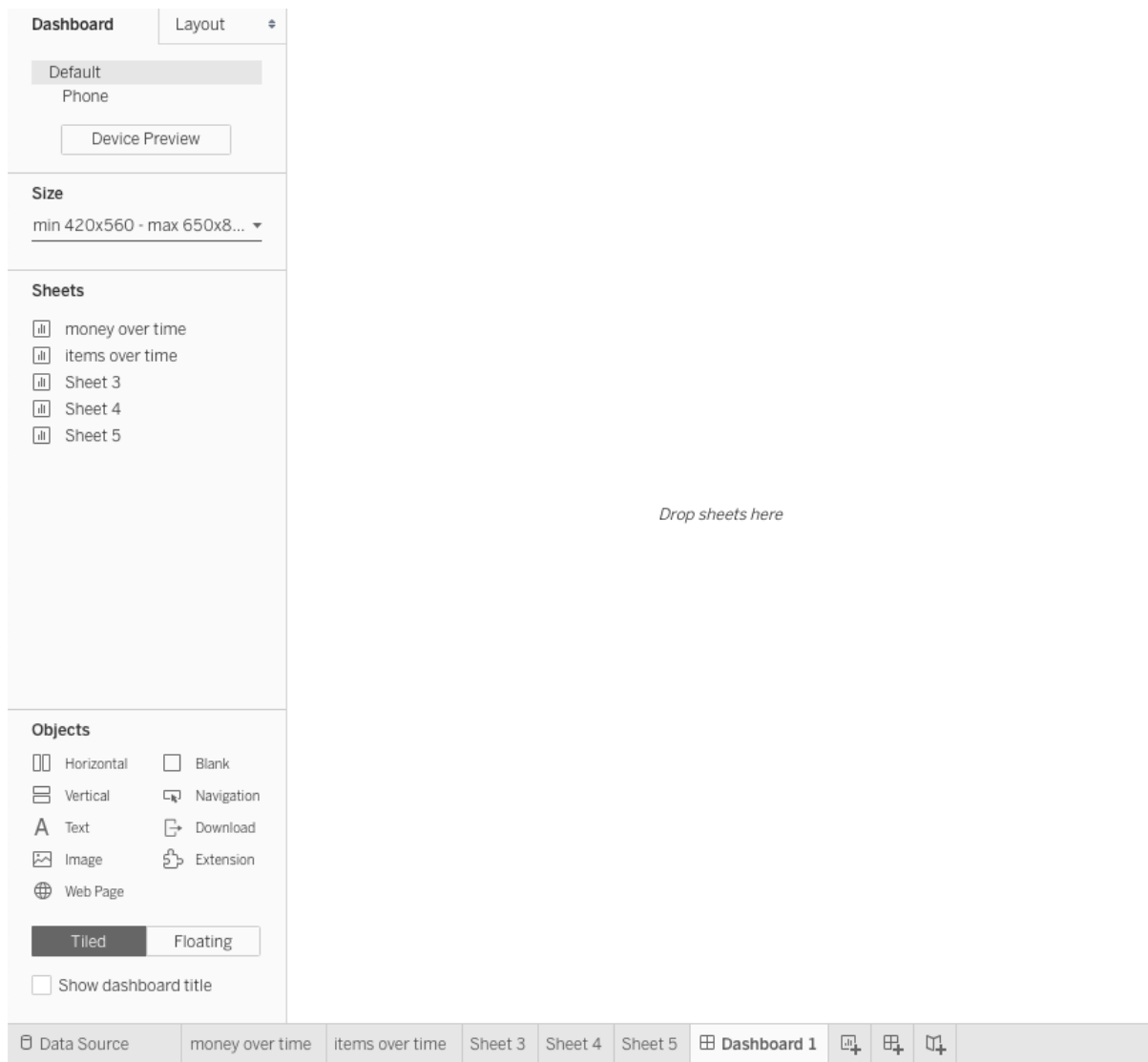


## > Интерфейс

Отдельные визуализации строятся на так называемых листах (worksheets/sheets). Новый лист можно добавить первой кнопкой с иконкой графика в нижней части экрана (каждый лист – отдельная визуализация). За создание дашборда отвечает следующая кнопка – *New Dashboard*. Истории – режим презентации; особенность состоит в том, что можно переключаться между различными состояниями визуализации.



Дашборд в Tableau состоит из ранее созданных графиков (sheets). Соответственно, чтобы добавить графики на дашборд, нужно перетащить на него нужные листы (sheet), полный список которых расположен в левой части экрана в *Sheets*. Изменить название можно двойным нажатием на него.



## > Основные понятия

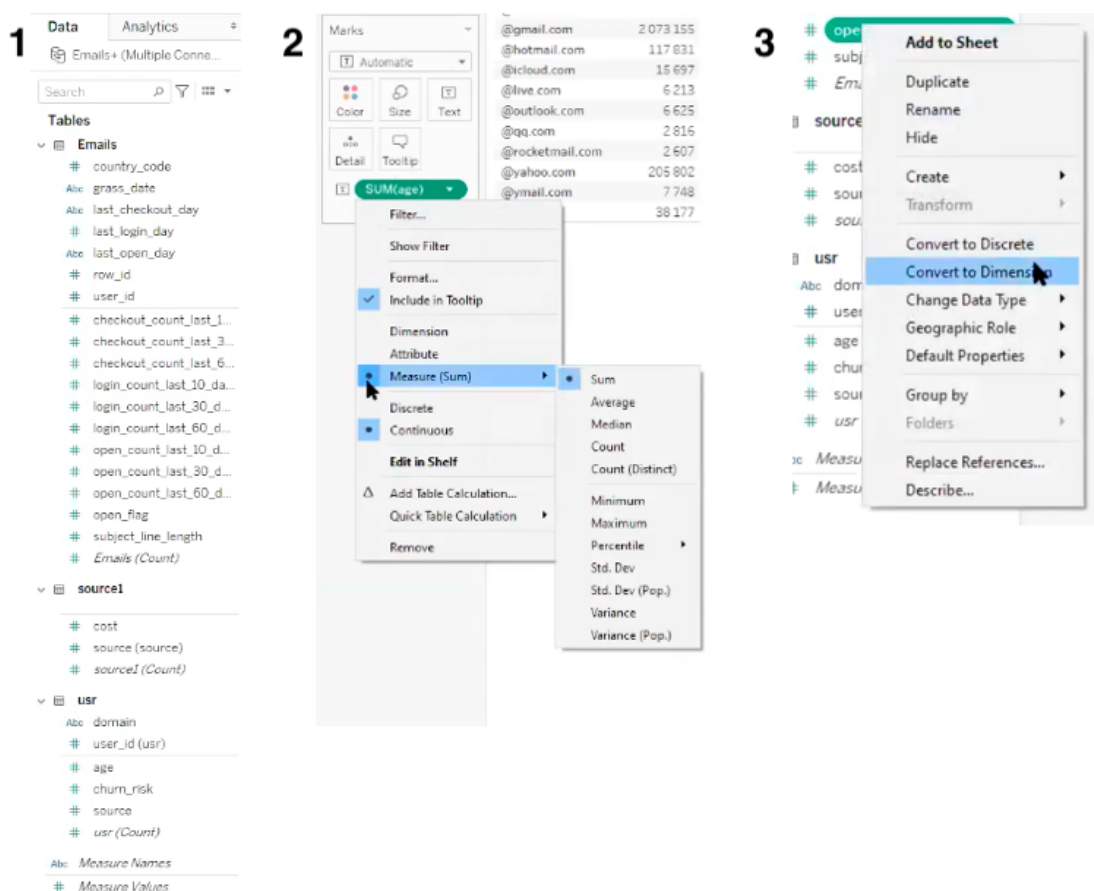
### Поля

Список подключенных таблиц и их колонок можно увидеть слева (1). Все они разделены по важному принципу – разделению на параметры (dimensions) и меры (measures).

- **Параметры** (dimensions) – срезы, которые делят наши данные по каким-то смысловым категориям. Например, по региону или продукту. Чаще всего – какие-то дискретные значения.
- **Меры** – метрики; то по каким параметрам мы хотим узнать что-то про измерения, т.е. то поле, которое мы агрегируем (например, число пользователей).

При добавлении измерения, которое мы хотели бы посмотреть, Tableau автоматически выбирает функцию агрегации. Изменить её можно нажав на поле (SUM(age)), далее Measure → нужная функция (2).

Дискретные значения помечены синим цветом, непрерывные – зелёным. Это важно, поскольку Tableau по-разному реагирует на эти поля в зависимости от комбинаций параметров и мер. Параметры также можно сделать мерами – для этого необходимо нажать на название параметра в списке слева и выбрать convert to measure. И наоборот – меру превратить в параметр с помощью convert to dimension (3).



Различия более наглядно:

	Discrete Создаёт разбивку/таблицу	Continuous Создаёт ось
Dimensions Чаше Дискретные  Всегда не агрегированы	<ul style="list-style-type: none"> <li>Пол человека (М/Ж)</li> <li>Возрастная группа (18-22, 23-30, 30-45)</li> <li>Оценка в школе (1,2,3,4,5)</li> <li>Регион</li> <li>Страна</li> <li>Тип пользователя</li> <li>Номер дома</li> <li>Дата (январь, февраль, март)</li> <li>Дата (2017, 2018, 2019)</li> </ul>	<ul style="list-style-type: none"> <li>Возраст (18, 20, 21.3...)</li> <li>Оценка в школе (1,2,3,6,4,5)</li> <li>Дата (январь 2017, февраль 2017, январь 2018)</li> <li>Дата (2017, 2018, 2019)</li> </ul>
Measures Чаше Непрерывные  Всегда агрегированы	<div style="text-align: center; font-size: 2em;">X</div>	<ul style="list-style-type: none"> <li>Возраст (2,5, 10, 15, 20, 20.4, 20.9)</li> <li>Трата (1500 р., 2340 р., 1205 р.)</li> <li>Температура (36.6°, 22°, 18°)</li> <li>Кол-во голосов (123, 342, 244)</li> </ul>

## Типы данных

В списке переменных (1) также можно увидеть значки, которые обозначают типы данных. Tableau определяет их автоматически. Для изменения типа переменной нужно нажать на значок и выбрать нужный. Наиболее популярные варианты:

- Number (decimal)
- Number (whole)
- Date & Time
- Date
- String

## Построение графика

В правой части экрана Tableau автоматически предлагает несколько видов графиков, наиболее подходящих для такого сочетания параметров и мер. Сами параметры и меры добавляются в раздел колонок (columns) и строк (rows). Если поменять их местами, то график повернется на 90 градусов.

Также кастомизировать график позволяют различные функции в поле marks (отметки). Отметки – способ отметить ту или иную переменную. Можно:

- Color – добавить цвет
- Size – размер (использовать в качестве размера)
- Label – лейблы (напр. добавить значения столбцов на график)
- Detail
- Tooltip

## > **Материалы**

1. [Презентация с лекции](#)
2. [Как сделать донат чарт](#)
3. [Поддержка Tableau](#)
4. [Бесплатное официальное обучение по Tableau](#)
5. [Виды визуализаций \(datavizproject\)](#).
6. [Виды визуализаций \(Tableau\)](#).
7. [Вебинар про Таблицы vs Графики](#)