

ОСНОВЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Принципы эффективной визуализации данных.

Типы визуализаций (гистограммы, столбчатые диаграммы, диаграммы рассеивания).

Типы визуализаций. Создание визуализаций с использованием Python: практический пример.

Что такое визуализация данных и почему это важно?

- **Визуализация данных** - это графическое представление информации.
- Визуализация данных и какое значение она имеет при практическом применении?



Понимание визуализации данных

- Визуализация данных переводит сложные наборы данных в визуальные форматы, которые легче понять человеческому мозгу.
- Это может включать различные визуальные инструменты, такие как:
 - **Диаграммы:** столбчатые диаграммы, линейные диаграммы, круговые диаграммы и т.д.
 - **Графы:** Рассеянные диаграммы, гистограммы и т.д.
 - **Карта:** Географические карты, тепловые карты и т.д.
 - **Доски управления:** Интерактивные платформы, объединяющие множество визуализаций.



Понимание визуализации данных

- Основная цель визуализации данных заключается в том, чтобы **сделать данные более доступными и легче интерпретируемыми**
- позволяя пользователям **быстро выявлять закономерности, тенденции и выбросы.**
- Это особенно важно в области больших данных, где **большой объем информации может сбить с толку без эффективных методов визуализации.**



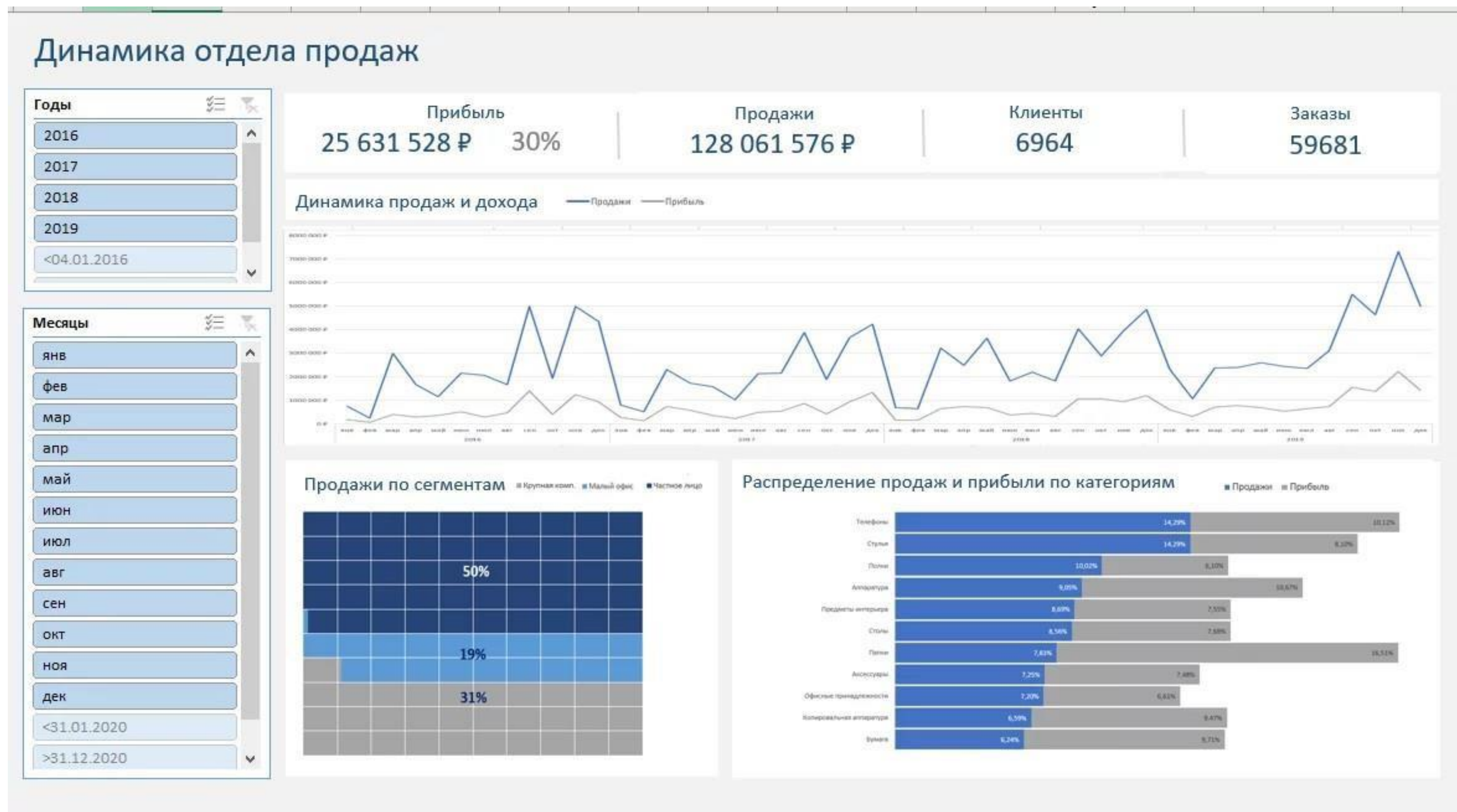
Почему важна визуализация данных?

- Давайте возьмем пример. Предположим, что вы собрали данные о прибыли компании с 2013 по 2023 годы и создали линейную диаграмму.
- Было бы очень легко увидеть, что линия постоянно растет с падением только в 2018 году. Таким образом, вы можете заметить в одно мгновение, что компания имела непрерывную прибыль в течение всех лет, за исключением убытка в 2018 году.
- Было бы нелегко получить эту информацию так быстро из таблицы данных. Это лишь одна демонстрация пользы визуализации данных.

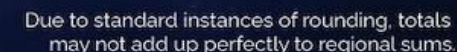
Табличные данные vs. Визуализированные данные

1	Name of Satellite Alternate Name	Current Official Name of Satell	Country/ Org o	Country of Operato	Operator/ Owner	Users	Purpose	Detailed Purpose	Class of Orbit	Type of Orbit	Longitude	Perigee (km)	Apogee (km)	Inclination (°)	Period (min)	Launch Date	Drift Rate	Power (watt)	Date of Launch	Expected Lifetime (yr)	Contractor	Country of Con
7084	STSat-2C	STSat-2C	South Kor	South Korea	Korea Aero	Government	Technology Development		LEO	Polar	0	298	1 482	80,28	102,77	93		160,00	30.01.2013	2	Korea Adv	South Korea
7096	SWARM-A	SWARM-A	ESA	ESA	European S	Government	Earth Observati	Earth Scienc	LEO	Polar	0	473	476	87,3	93,9	468	369	608,00	22.11.2013	4	Astrium	France
7097	SWARM-B	SWARM-B	ESA	ESA	European S	Government	Earth Observati	Earth Scienc	LEO	Polar	0	473	476	87,3	93,9	468	369	608,00	22.11.2013	4	Astrium	France
7098	SWARM-C	SWARM-C	ESA	ESA	European S	Government	Earth Observati	Earth Scienc	LEO	Polar	0	486	496	87,5	94	468	369	608,00	22.11.2013	4	Astrium	France
7099	Swift	Swift	USA	USA/United	Goddard Sp	Government	Space Science		LEO	Non-Polar Incline	0	551	576	20,6	96,3	1 463		1 040,00	20.11.2004	7	Spectrum A	USA
7101	Syracuse 3A (Syste	Syracuse 3A	France	France	Defense M	Military	Communications		GEO		47,05	35 778	35 792	0	1436,06	3 725	1 650	5 640,00	13.10.2005	12	Alcatel Alen	France/Italy
7104	TacSat 4	TacSat 4	USA	USA	Naval Rese	Military	Technology Development		Elliptical	Molniya	0	658	11 836	63,64	235,16	450		1 000,00	27.09.2011		Naval Rese	USA
7112	Tango (part of Clus	Tango	ESA	ESA	European S	Government	Space Science		Elliptical	Deep Highly Ecc	0	17 313	120 961	90,7	3442	1 200	478	220,00	09.08.2000	4	Dornier Syst	International
7117	TDRS-10 (Tracking	TDRS-10	USA	USA	National Ae	Government	Communications		GEO		-40,8	35 768	35 802	1,64	1436,04	3 180	1 600	2,300 (EOL)	04.12.2002	11	Boeing Sate	USA
7118	TDRS-11 (Tracking	TDRS-11	USA	USA	National Ae	Government	Communications		GEO		-171	35 732	35 844	7	1436,18	3 454	1 600	2,300 (EOL)	31.01.2013	15	Boeing Sate	USA
7119	TDRS-12 (Tracking	TDRS-12	USA	USA	National Ae	Government	Communications		GEO		-41	35 780	35 785	6,99	1435,88	3 454	1 600	3,500 (EOL)	23.01.2014	15	Boeing Sate	USA
7120	TDRS-13 (Tracking	TDRS-13	USA	USA	National Ae	Government	Communications		GEO		-150	35 755	35 813	0	1436	3 452	1 600	3,500 (EOL)	18.08.2017	15	Boeing Sate	USA
7121	TDRS-3 (Tracking a	TDRS-3	USA	USA	National Ae	Government	Communications		GEO		-49	35 693	35 878	11,53	1436,06	3 180	1 600	1 800,00	29.09.1988	10	TRW Defen	USA
7122	TDRS-5 (Tracking a	TDRS-5	USA	USA	SES S.A.	Government	Communications		GEO		-174	35 773	35 796	9,4	1436,99	3 180	1 600	1 800,00	02.08.1991	10	TRW Defen	USA
7123	TDRS-6 (Tracking a	TDRS-6	USA	USA	National Ae	Government	Communications		GEO		-171	35 773	35 787	8,74	1435,7	3 180	1 600	1 800,00	13.01.1993	10	TRW Defen	USA
7124	TDRS-7 (Tracking a	TDRS-7	USA	USA	National Ae	Government	Communications		GEO		-151	35 767	35 803	10,6	1436,04	3 180	1 600	1 800,00	13.07.1995	10	TRW Defen	USA
7125	TDRS-8 (Tracking a	TDRS-8	USA	USA	National Ae	Government	Communications		GEO		88,76	35 768	35 808	0,9	1436,2	3 180	1 600	2,300 (EOL)	30.06.2000	11	Boeing Sate	USA
7126	TDRS-9 (Tracking a	TDRS-9	USA	USA	National Ae	Government	Communications		GEO		-62,4	35 768	35 809	0,02	1436,11	3 180	1 600	2,300 (EOL)	08.03.2002	11	Boeing Sate	USA
7132	Telkom 3S	Telkom 3S	Indonesia	Indonesia	PT Telkom	Commercial	Communications		GEO		118	35 780	35 795	0	1436,1	3 550		6 400,00	14.02.2017	18	Thales Alen	France/Italy
7135	Telstar 14R (Estrela	Telstar 14R	Brazil	Canada	Telesat Can	Commercial	Communications		GEO		-63	35 783	35 791	0,02	1436,1	4 970	2 150	12 000,00	20.05.2011	15	Space Syste	USA
7143	TET-1 (Technologie	TET-1	Germany	Germany	German Ae	Commercial	Technology Development		LEO	Sun-Synchronou	0	501	505	97,5	94,7	120		20,00	22.07.2012	1	Kayser-Thre	Germany
7153	Thaicom-4 (Ipstar	Thaicom-4	NR	Thailand	Thaicom Pu	Commercial	Communications		GEO		119,5	35 775	35 798	0	1436,1	6 505		14 000,00	11.08.2005	12	Space Syste	USA
7157	THEMIS A (Time Hi	THEMIS A	USA	Multination	National Ae	Government	Space Science		Elliptical		0	461	87 304	15,7	1868,98	126	77	40,00	17.02.2007	2	Swales Aerc	USA
7158	THEMIS D (Time Hi	THEMIS D	USA	Multination	National Ae	Government	Space Science		Elliptical		0	467	87 260	15,7	1867,91	126	77	40,00	17.02.2007	2	Swales Aerc	USA
7159	THEMIS E (Time Hi	THEMIS E	USA	Multination	National Ae	Government	Space Science		Elliptical		0	474	87 526	15,7	1875,53	126	77	40,00	17.02.2007	2	Swales Aerc	USA
7160	THEOS (Thailand E	Theos	Thailand	Thailand	Geo-Inform	Government	Earth Observati	Optical Imag	LEO	Sun-Synchronou	0	824	826	98,7	101,4	750	670	840,00	01.10.2008	5	EADS Astriu	France
7162	Thor-6	Thor-6	Norway	Norway	Telenor Sat	Commercial	Communications		GEO		-1	35 779	35 794	0,02	1436,1	3 050		5 900,00	29.10.2009	15	Thales Alen	France/Italy
7164	Thuraya 2	Thuraya 2	United Ar	United Arab	Thuraya Sa	Commercial	Communications		GEO		44,13	35 764	35 807	2,99	1436,07	3 200	3 200	11 000,00	10.06.2003	12	Hughes Spa	USA
7233	TIMED (Thermosph	TIMED	USA	USA	National Ae	Government	Space Science		LEO	Polar	0	612	613	74,1	97	587		406,00	07.12.2001	2	Applied Phy	USA
7252	TUGSat-1 (Technis	TUGSat-1	Austria	Austria	Technical U	Civil	Technology Development		LEO	Sun-Synchronou	0	770	784	98,6	100,3	14	7	5,40	25.02.2013		University o	Canada
7254	Turksat 3A	Turksat 3A	Turkey	Turkey	Turksat	Commercial	Communications		GEO		42,04	35 778	35 793	0,01	1436,06	3 100	1 272	8 100,00	12.06.2008		Thales Alen	France
7272	UFO-10 (USA 146,	USA 146	USA	USA	US Navy	Military	Communications		GEO		72,67	35 738	35 834	1,97	1436,1	3 200	1 540	3,800 (EOL)	24.11.1999	4	Boeing Sate	USA
7273	UFO-11 (USA 174)	USA 174	USA	USA	US Navy	Military	Communications		GEO		71,44	35 786	35 787	2,16	1436,1	3 200	1 545	2,800 (EOL)	18.12.2003	4	Boeing Sate	USA
7274	UFO-2 (USA 95) "U	USA 95	USA	USA	US Navy	Military	Communications		GEO		28,79	35 766	35 807	8,27	1436,12	3 200	1 545	2,800 (EOL)	03.09.1993		Boeing Sate	USA
7275	UFO-4 (USA 108, U	USA 108	USA	USA	US Navy	Military	Communications		GEO		-177	35 763	35 775	3,34	1435,22	3 200	1 545	2,800 (EOL)	29.01.1995	4	Boeing Sate	USA
7276	UFO-8 (USA 138, U	USA 138	USA	USA	US Navy	Military	Communications		GEO		172,3	35 772	35 794	5,24	1435,94	3 206	1 545	3,800 (EOL)	16.03.1998	4	Boeing Sate	USA
7287	UNIFORM 1 (UNiv	UNIFORM 1	Japan	Japan	Wakayama	Civil	Earth Observati	Optical Imag	LEO	Sun-Synchronou	0	622	629	97,87	97,2	50		140,00	24.05.2014		Wakayama	Japan
7321	VIREO	VIREO	NR (9/23)	Hungary	C3S LLC	Commercial	Technology Demonstration		LEO	Sun-Synchronou	0	487	507		97,4	95	4	4,00	15.04.2023		C3S	Hungary
7335	Wideband Global S	USA 195	USA	USA	Military Sat	Military	Communications		GEO		179,9	35 783	35 790	0	1436,1	5 900	3 450	13 000,00	11.10.2007	14	Boeing Inte	USA
7336	Wideband Global S	USA 291	USA	USA	Military Sat	Military	Communications		GEO		-52	35 765	35 795	0	1436,1	5 990	3 450	13 000,00	16.03.2019	14	Boeing Sate	USA
7337	Wideband Global S	USA 204	USA	USA	Military Sat	Military	Communications		GEO		60,18	35 786	35 787	0,03	1436,12	5 987	3 450	13 000,00	04.04.2009	14	Boeing Sate	USA
7338	Wideband Global S	USA 211	USA	USA	Military Sat	Military	Communications		GEO		-12	35 785	35 786	0	1436,1	5 990	3 450	13 000,00	06.12.2009	14	Boeing Sate	USA
7339	Wideband Global S	USA 233	USA	USA	Military Sat	Military	Communications		GEO		-125	35 785	35 786	0,01	1436,1	5 990	3 450	13 000,00	20.01.2012	14	Boeing Sate	USA
7340	Wideband Global S	USA 243	USA	USA	Military Sat	Military	Communications		GEO		-52,5	35 739	35 784	0	1436,1	5 990	3 450	13 000,00	25.05.2013	14	Boeing Sate	USA
7341	Wideband Global S	USA 244	USA	USA	Military Sat	Military	Communications		GEO		-135	35 772	35 801	0	1436,1	5 990	3 450	13 000,00	08.08.2013	14	Boeing Sate	USA
7342	Wideband Global S	USA 263	USA	USA	Military Sat	Military	Communications		GEO		175	35 772	35 801	0	1436,1	5 990	3 450	13 000,00	24.07.2015	14	Boeing Sate	USA
7343	Wideband Global S	USA 272	USA	USA	Military Sat	Military	Communications		GEO		175	35 772	35 801	0	1436,1	5 990	3 450	13 000,00	07.12.2015	14	Boeing Sate	USA

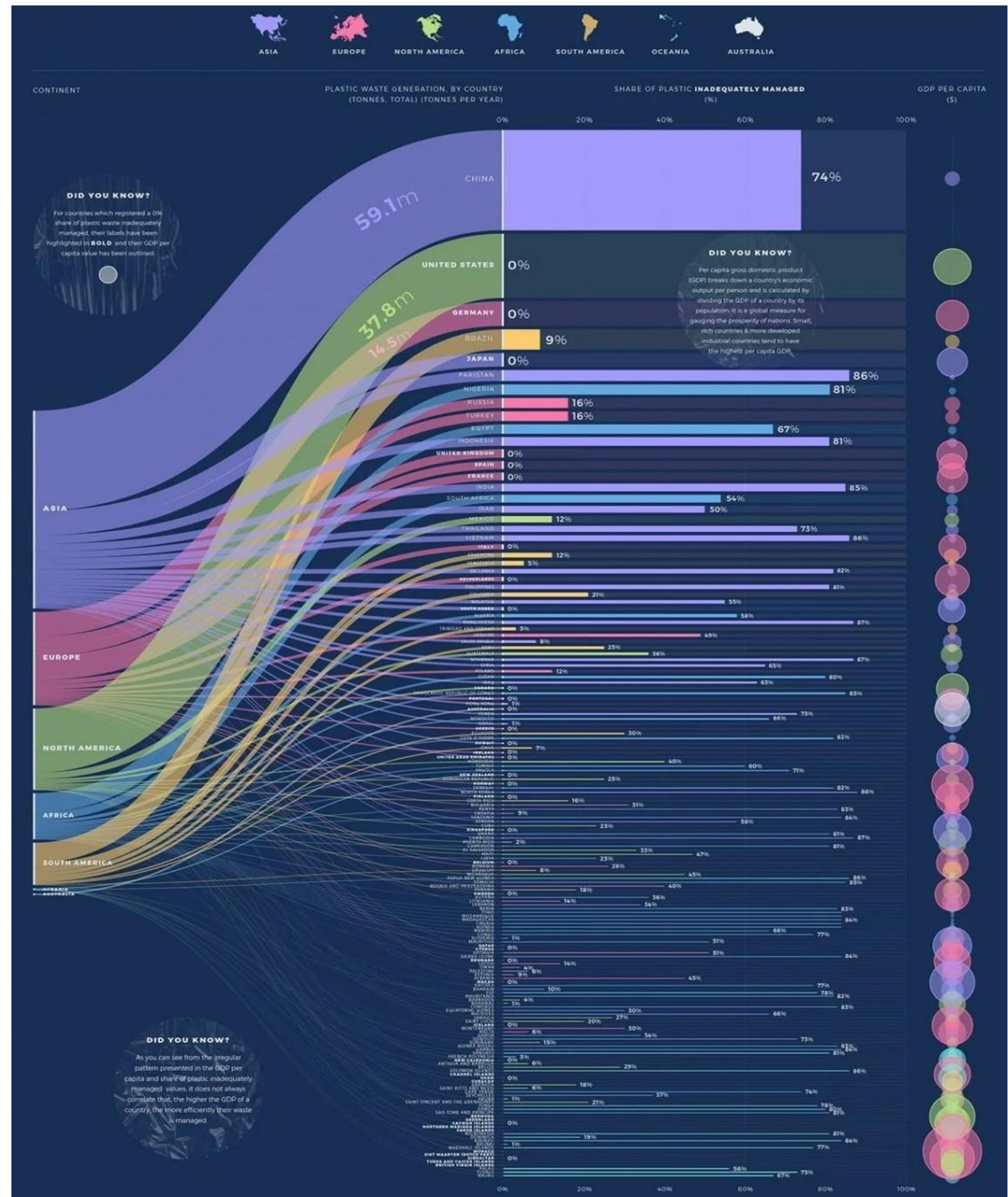
Табличные данные vs. Визуализированные данные



Население Земли

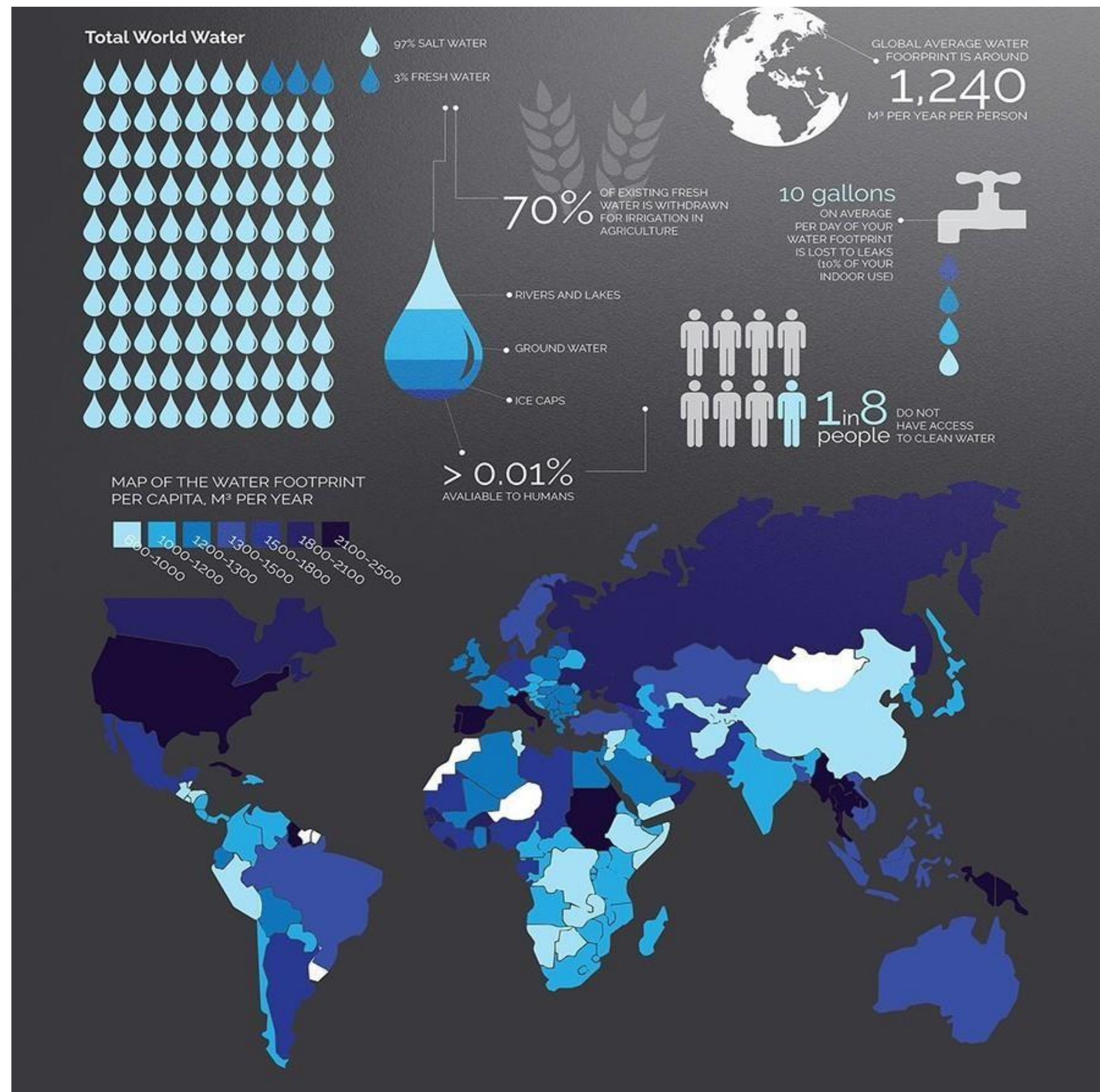


Загрязнение пластиковыми отходами



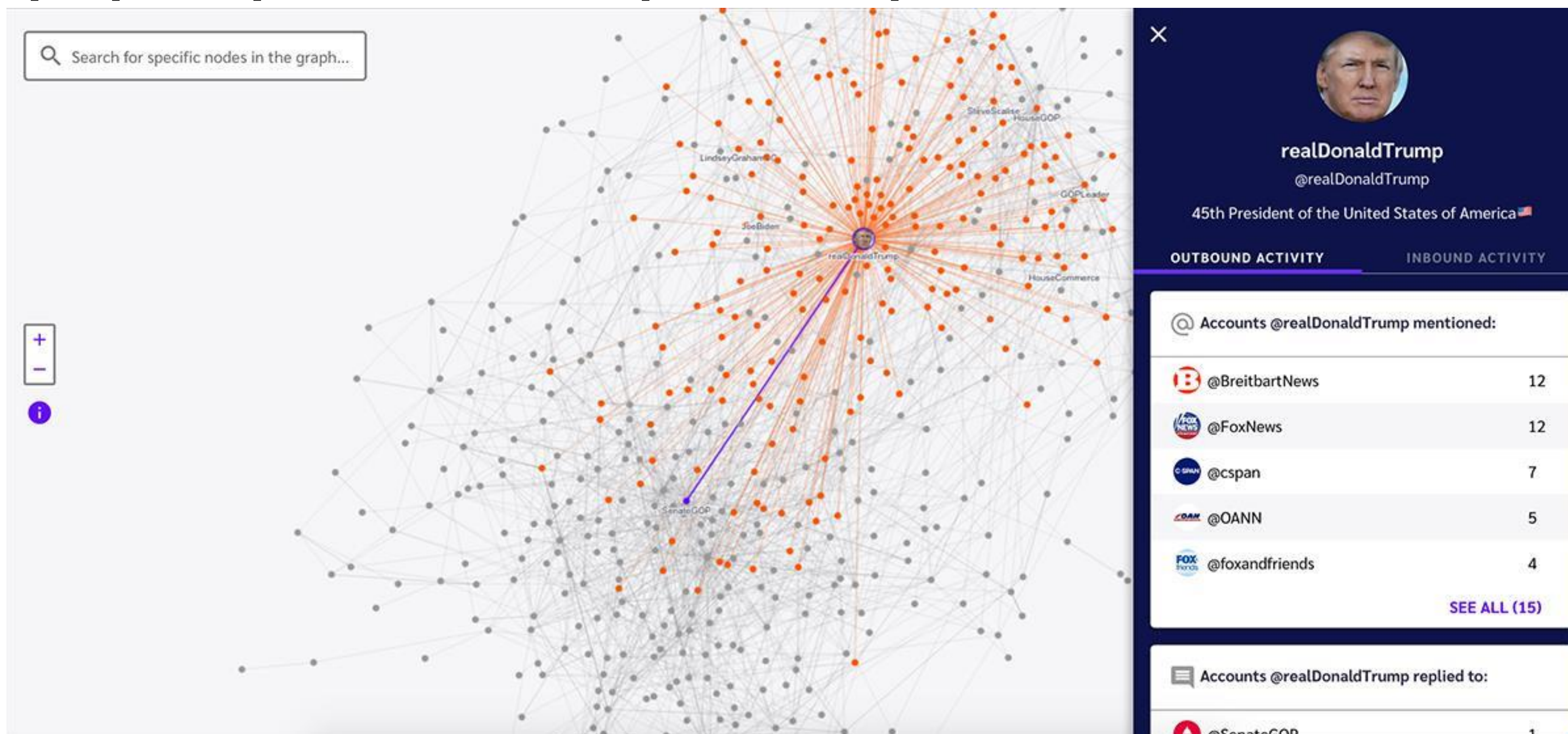
Примеры эффективных визуализации

Потребление воды



Примеры эффективных визуализации

График продвижения предвыборной кампании в Twitter



Эффективность визуализации данных

- Создание визуализации данных требует определенных усилий и времени.
- Но они с лихвой окупаются, потому что у такого формата подачи данных есть неоспоримые достоинства.



Эффективность визуализации данных

- Информацию легко и удобно анализировать. Множество таблиц просто не нужно
- Пользователь получает целостную картину сразу по всем значимым показателям
- Грамотная визуализация не только упрощает восприятие, но и повышает уровень доверия к данным
- Позволяет посмотреть на данные в разных разрезах и получить инсайты, незаметные в простом отчете
- Обсуждение показателей внутри компании становится более конструктивным, потому что отчетность понятна и наглядна для всех



Основные принципы визуализации

- **Четкость и простота**
 - Минимизация лишней информации
 - Избегание избыточных деталей и "шумных" графиков.
- **Правильный выбор типов графиков**
 - Когда использовать гистограммы, линейные графики, круговые диаграммы и др.
 - Ошибки при выборе типов визуализации.

Основные принципы визуализации

- **Использование цвета**
 - Как цвета влияют на восприятие данных
 - Принципы цветовой гармонии
- **Контекст и аннотации**
 - Подписи, легенды, заголовки
 - Как правильно представлять данные для разных аудиторий

Выбор диаграммы в соответствии с задачей



Сравнение

- Сходство и различие между переменными в один момент времени.
- Например, сравнение результатов маркетинговых кампаний, план-факт продаж.



Линейчатая диаграмма (Bar Chart)

Также известна как столбчатая или полосчатая диаграмма, отображающая информацию в виде баров (столбцов). Длина столбца – значение параметра



Линейчатая диаграмма с группировкой

Используется для сравнения двух или нескольких наборов данных. Каждый набор данных выделен своим цветом.



Вертикальная диаграмма Водопад

Используется для отображения изменений значения относительно предыдущего. Верхний – исходное значение, нижний – итоговое.



Лолипоп

Значение параметра отображено длиной линий, на верхушке каждой добавлен кружок. Вид диаграммы позволяет добавлять сноски и пояснения.

Распределение

- Деление на категории, частотность в пределах интервала, группировка по признаку.
- Например, распределение населения по возрастам, продаж по регионам, процентное соотношение успешных и провальных событий.



Гистограмма

Визуализирует распределение данных в рамках непрерывного интервала или ограниченного периода времени.



Боксплот

Диаграмма размаха, «ящик с усами». Показывает статистические показатели: медиану, верхний и нижний квартили, минимумы и максимумы, выбросы.



Точечная карта

Точечная карта плотности распределения. Отображает географические паттерны или распределение данных по географическим регионам.



График плотности

Демонстрирует распределение данных за интервал времени. Отображает гладкое распределение за счет сглаживания изменений параметров.

Структура

- Соотношение части к целому, состав, деление категорий на подкатегории и их отношение к общей величине.
- Например, доля филиалов в общей выручке, влияние увеличения продаж и сокращения расходов на рост прибыли.



Круговая диаграмма (Pie chart)

Отражает соотношение категорий к целому при помощи сегментов окружности. Вся окружность означает общую сумму, 100%.



Линейчатая диаграмма с накоплением

Показывает, как крупная категория делится на более мелкие, и какое отношение имеет каждая часть в рамках общей величины.



Горизонтальная диаграмма Водопад

Показывает процентное соотношение подкатегорий и общего показателя, изменение частей от целого во времени.



Древовидная карта (Treemap)

Отображает иерархические данные, полезна при необходимости визуализировать большие объемы данных.

Корреляция

- Взаимосвязь между переменными, которую можно описать как "когда изменяется переменная X, изменяется и переменная Y".
- Например, сегментация клиентов по частоте обращений, возрасту и покупательской способности.



Пузырьковая диаграмма

Отображает трехмерные данные в двумерном пространстве. Размер каждого пузырька является третьим измерением.



Диаграмма рассеяния

Демонстрирует, существуют ли связи между двумя переменными. Степень взаимосвязи определяется плотностью расположения точек.



Зеркальная линейчатая диаграмма

Два набора данных располагаются по разным сторонам оси для удобства сравнения и поиска взаимосвязей.



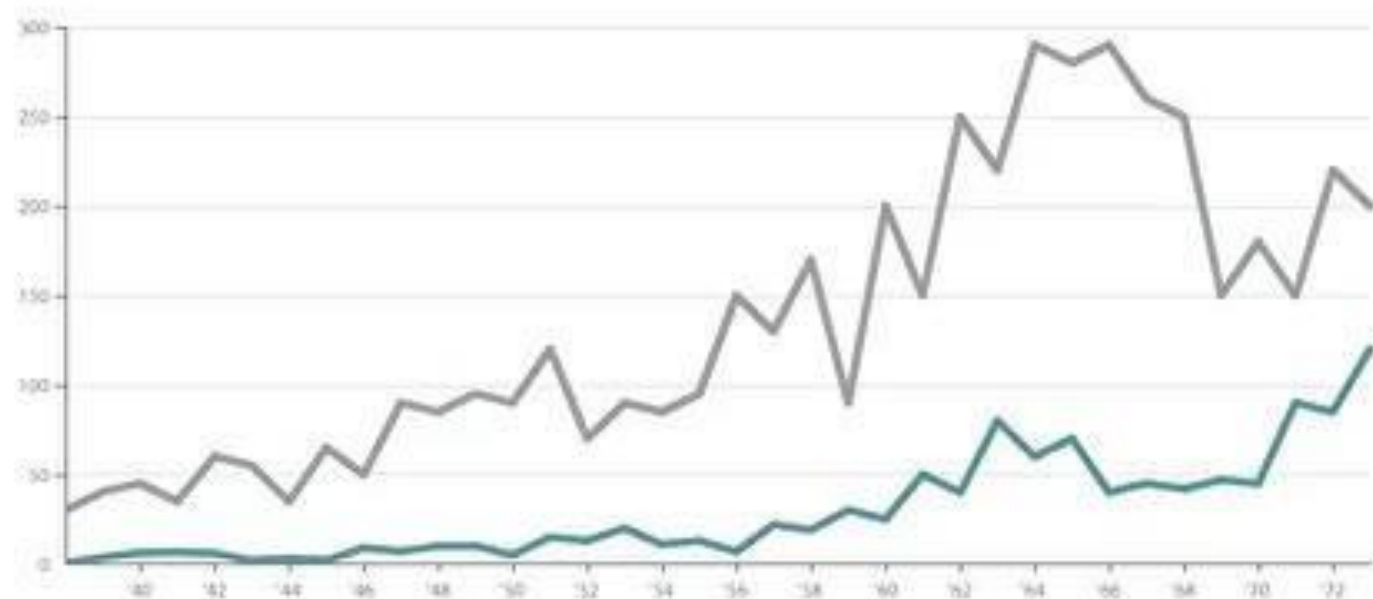
Тепловая карта

Удобна для демонстрации вариативности переменных, выявления паттернов, отображения взаимосвязей между переменными.

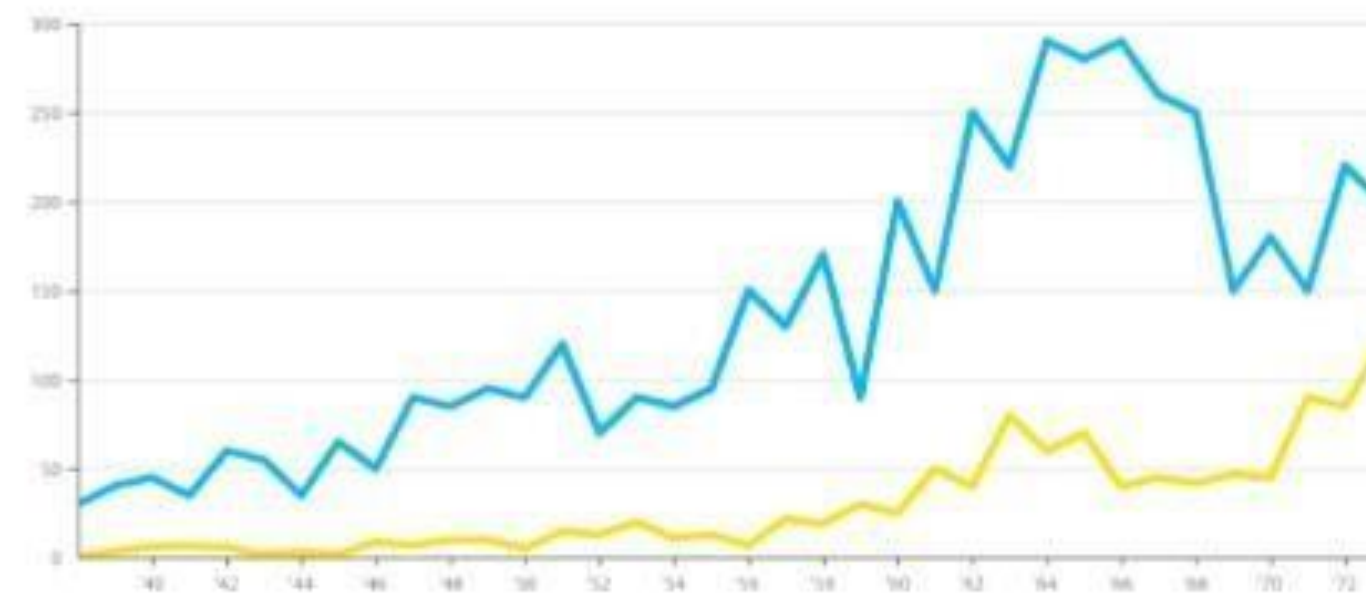
Правильно выбранная визуализация помогает представить данные более понятно и наглядно, обнаружить скрытые тренды или аномалии в данных.

Правильное использование цветов

- Цветовая палитра для визуализаций способна как помочь пользователю быстрее оценить ключевые показатели, так и создать хаос в данных.



✗ Сложно сравнить показатели

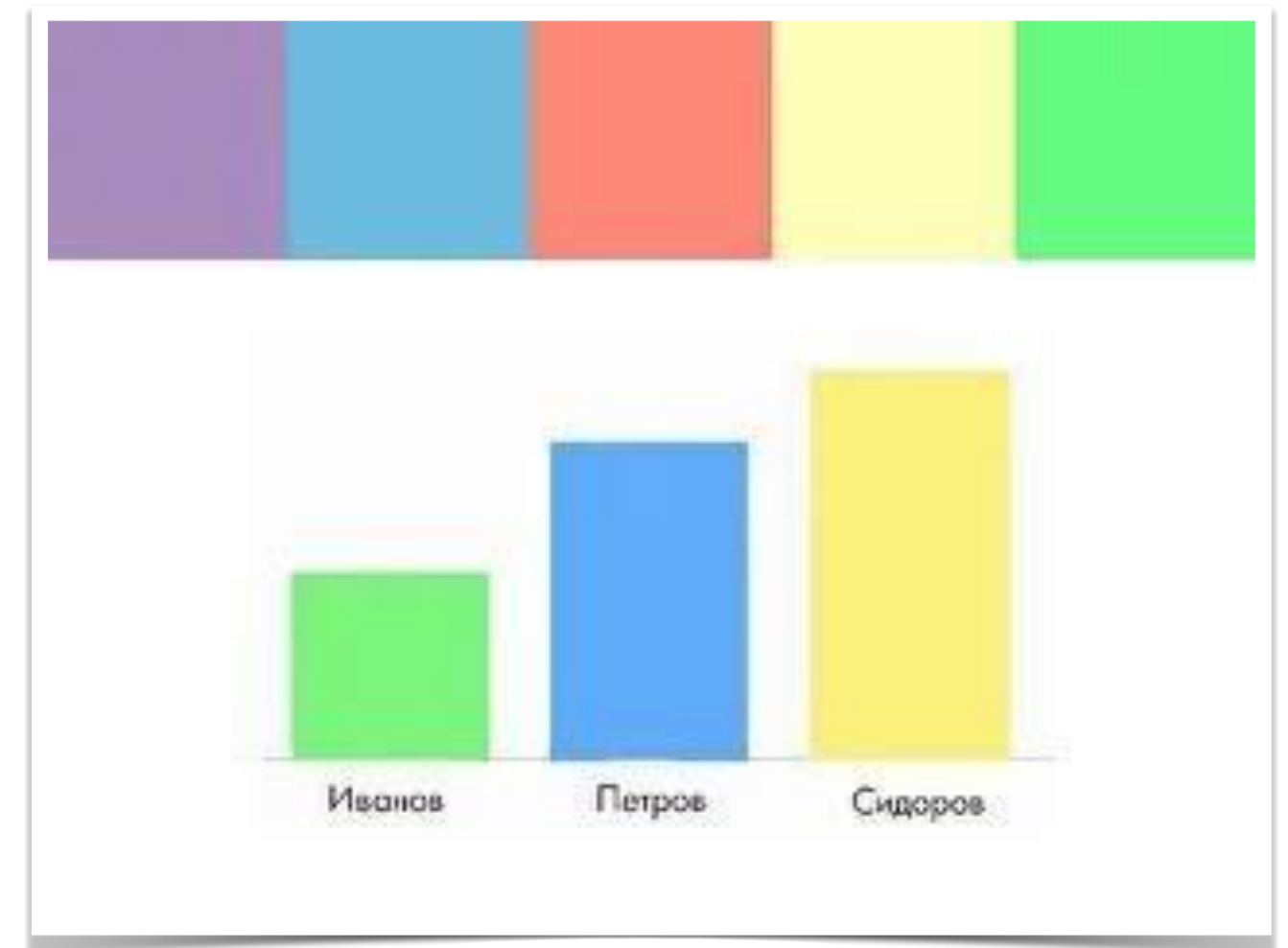


✓ Проще сравнить благодаря контрастным цветам

Цветовая палитра зависит от типа данных

Категориальная (качественная) палитра

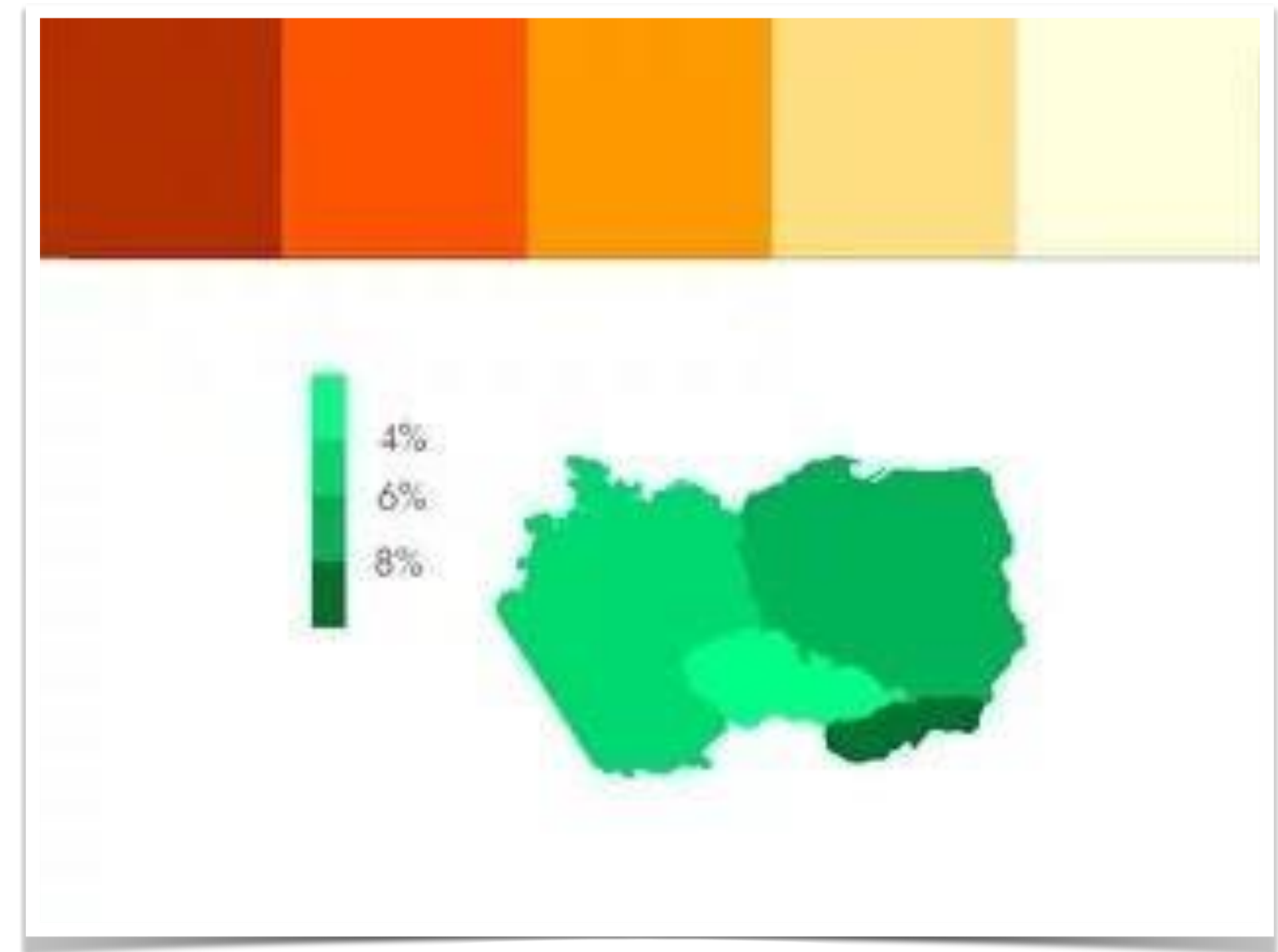
- Палитра состоит из разных цветов, которые не имеют упорядоченного или градиентного значения, с близкой яркостью и насыщенностью.
- Категориальная цветовая палитра предназначена для представления категорий или классов данных без какого-либо внутреннего порядка или значений между категориями.
- Она используется для отображения дискретных категорий, таких как различные группы товаров, типы событий, категории пользователей и т. д.



Цветовая палитра зависит от типа данных

Последовательная палитра

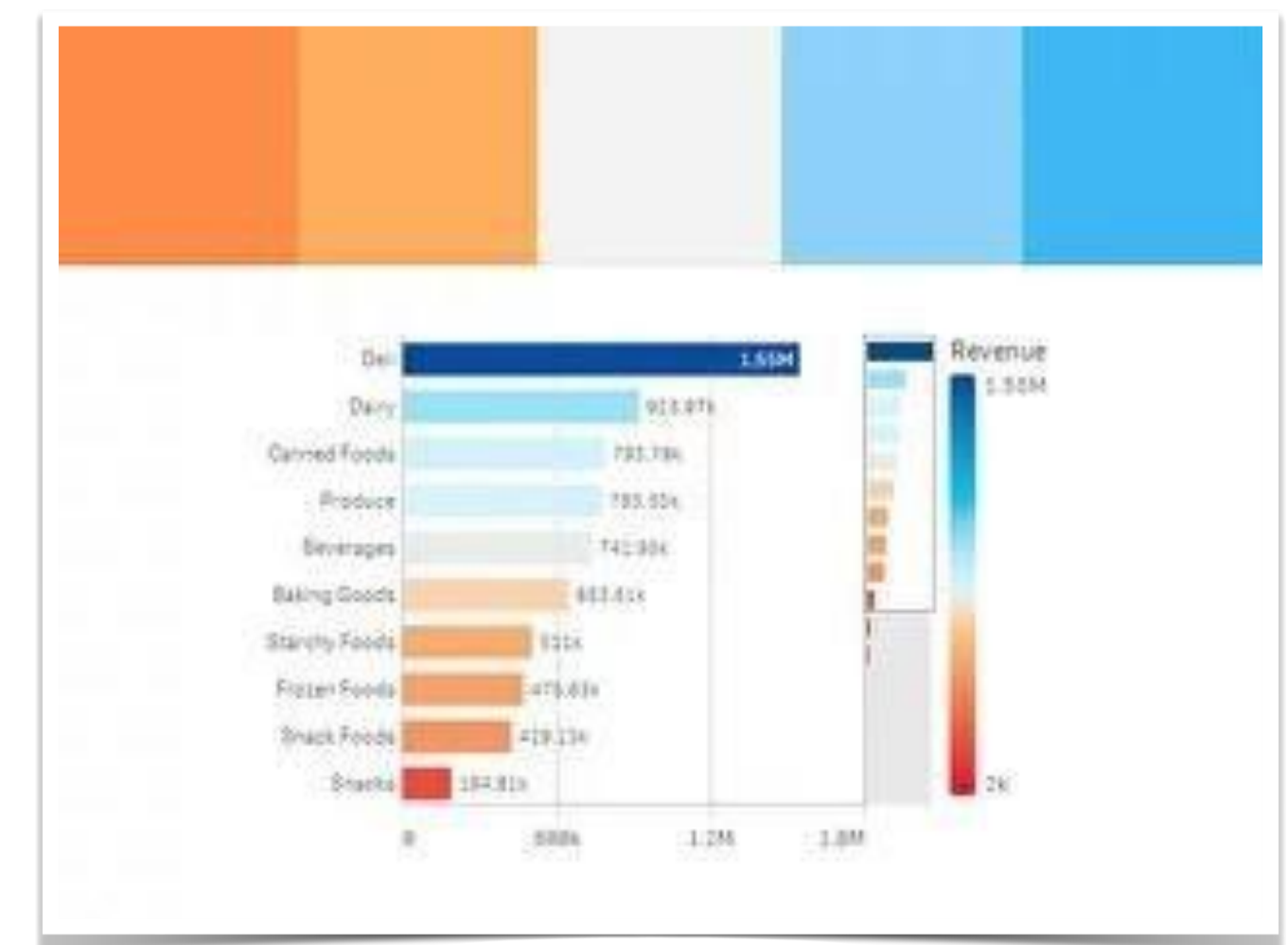
- Градиент из одного цвета разной насыщенности.
- Палитру используют для данных, где присутствует упорядоченная структура, явная прогрессия или различные уровни значений.
- Например, такая палитра идеально подходит для отображения температурных карт, рейтингов, показателей продаж, или любых данных, которые имеют явное распределение от низких к высоким значениям или наоборот.



Цветовая палитра зависит от типа данных

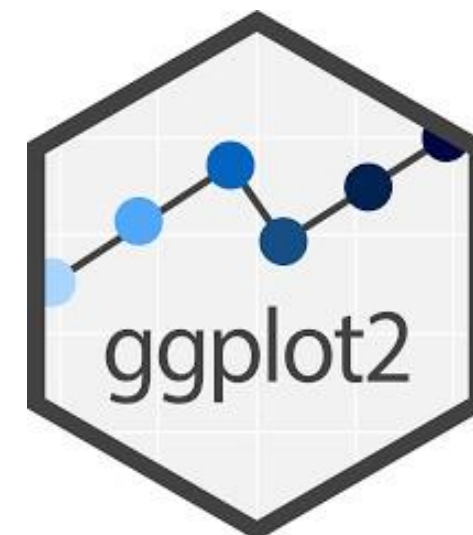
Расходящаяся палитра

- Комбинация последовательных палитр с разрывом, диапазон значений с определенной серединой, от которой цвета уходят в разные направления.
- Расходящаяся палитра подходит для визуализации данных, которые имеют явное различие между низкими и высокими значениями, с выделенной областью средних значений или уровнем концентрации.
- Это может быть полезно при отображении данных о средней температуре, статистике доходов или других показателях, где центральное значение играет важную роль.



Выбор инструментов визуализации

- Excel
- Google Sheets
- Power BI
- Tableau
- Python (Matplotlib, Seaborn)
- R (ggplot2)



Примеры на Python

Гистограмма с Matplotlib

```
import matplotlib.pyplot as plt

import numpy as np

data = np.random.randn(1000)

plt.hist(data, bins=30, edgecolor='black')

plt.title('Распределение данных')

plt.xlabel('Значение')

plt.ylabel('Частота')

plt.show()
```

Примеры на Python

Линейный график с Seaborn

```
import seaborn as sns

import pandas as pd

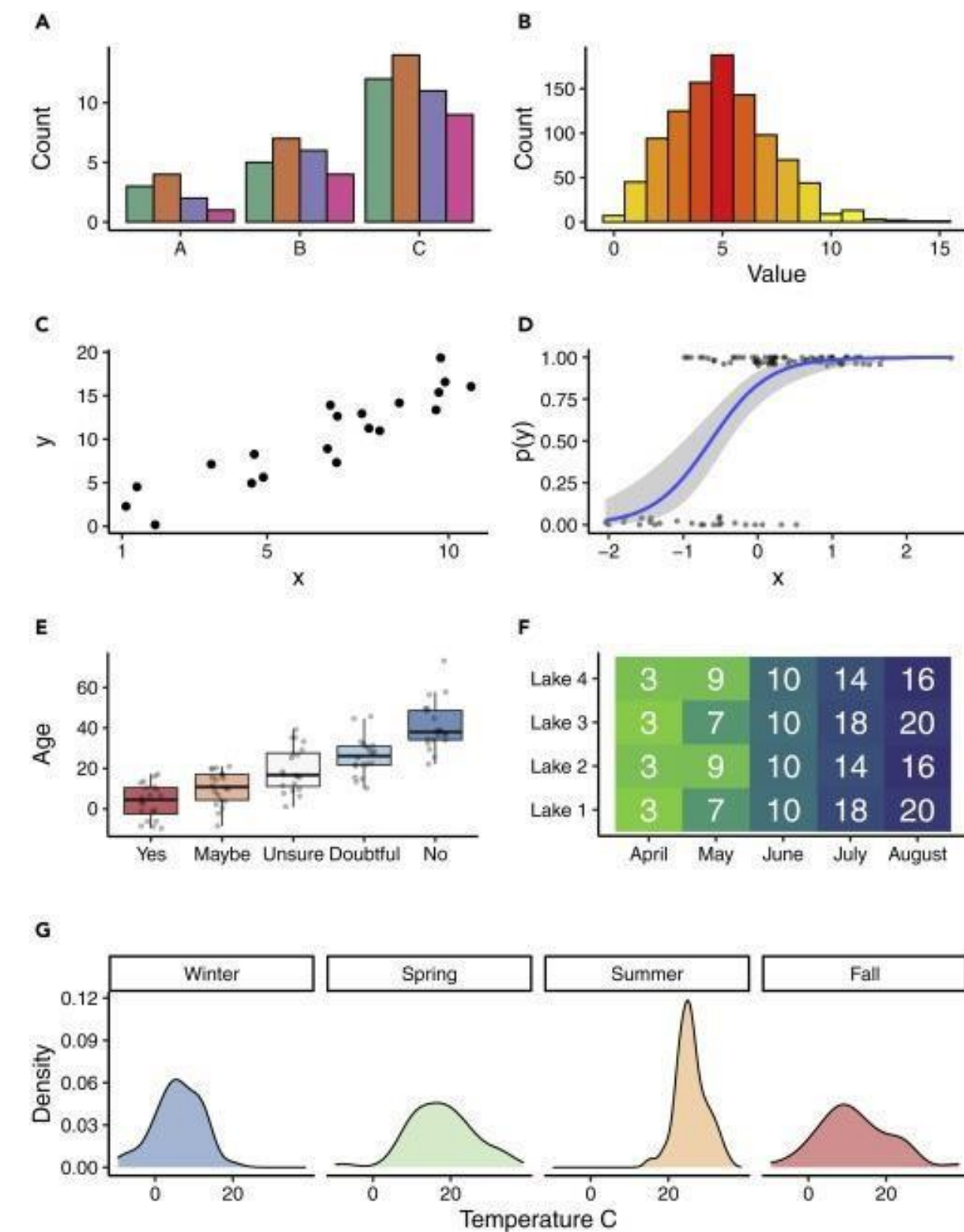
df = pd.DataFrame({
    'Год': [2018, 2019, 2020, 2021, 2022],
    'Продажи': [200, 240, 300, 500, 700]
})

sns.lineplot(x='Год', y='Продажи', data=df, marker='o')
plt.title('Динамика продаж по годам')

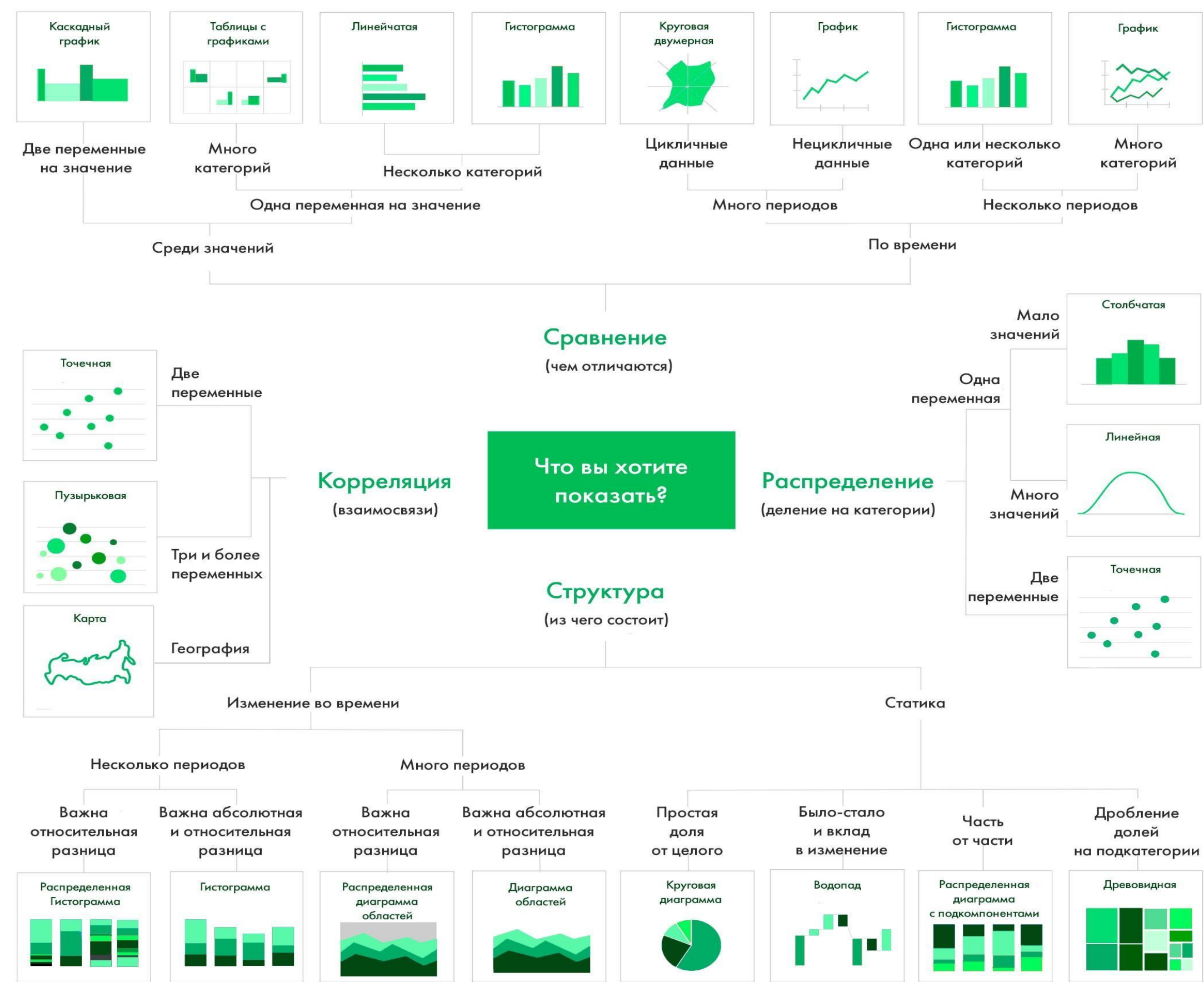
plt.show()
```


Как создать хорошие визуализации?

- **Выбрать диаграмму**, правильно отображающую конкретный тип данных
- **Использовать цвета** так, чтобы они помогли пользователю быстрее оценить ключевые показатели
- **Расставить акценты** в формах, размерах, с помощью насыщенности цветов, которые подчеркнут самое важное
- **Выбрать такой дизайн визуализаций**, который не будет препятствовать пониманию связи в данных или искажать их



Выбор диаграммы в соответствии с задачей



Выбор цвета

Чем меньше цветов, тем проще найти их на дашборде

- Для создания визуализаций лучше использовать не больше 5-6 цветов.
- Диаграмма будет выглядеть более профессионально, если использовать в ней только несколько оттенков и комплементарных цветов.



✗ Сложно сфокусироваться

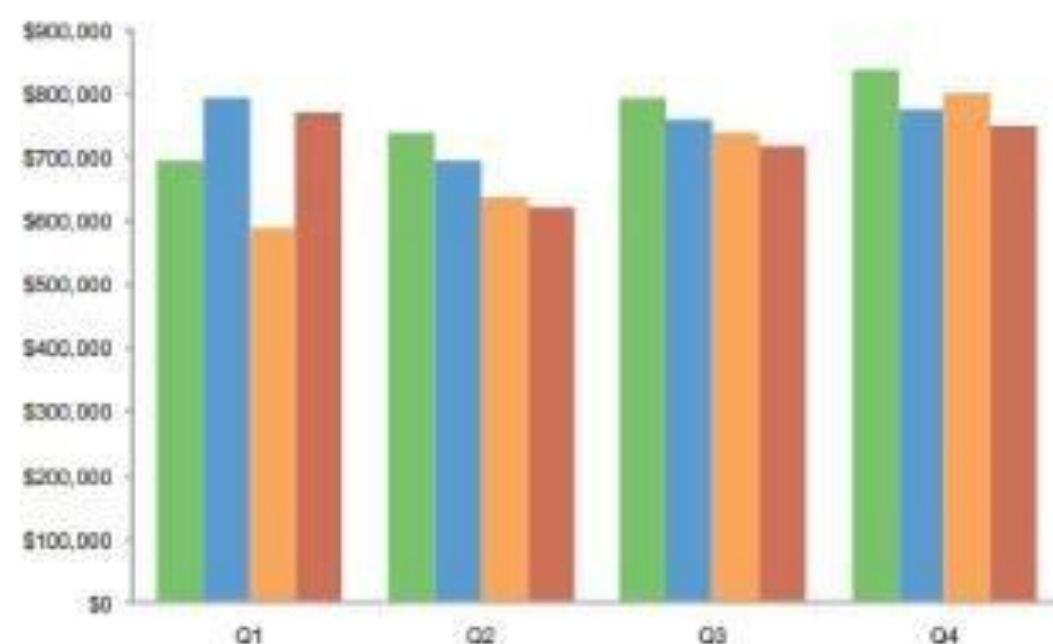


✓ Понятнее

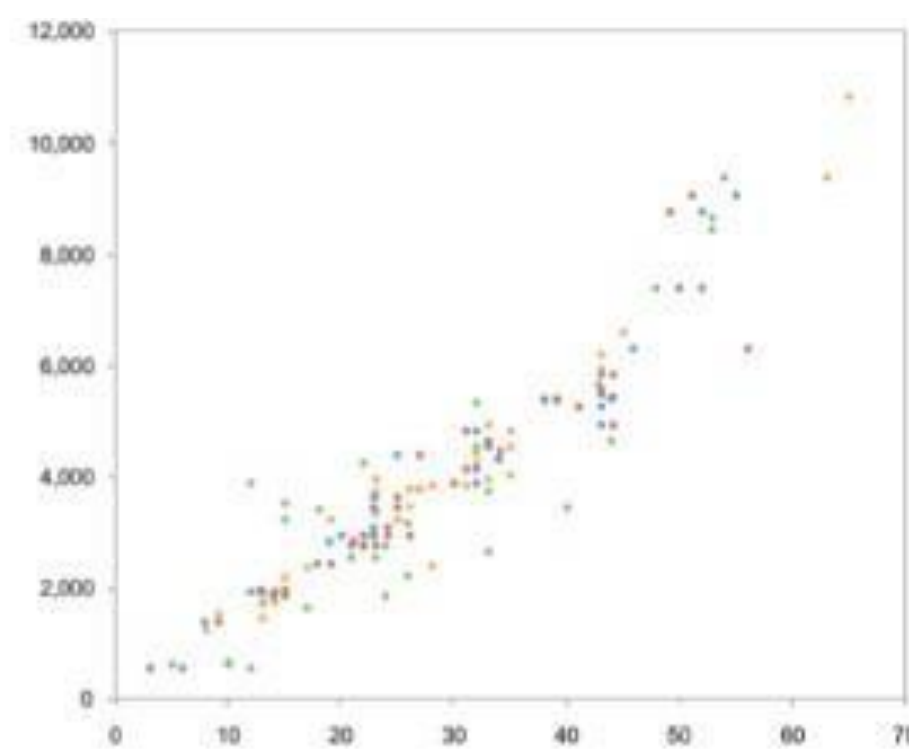
Выбор цвета

Разные цвета для разных диаграмм

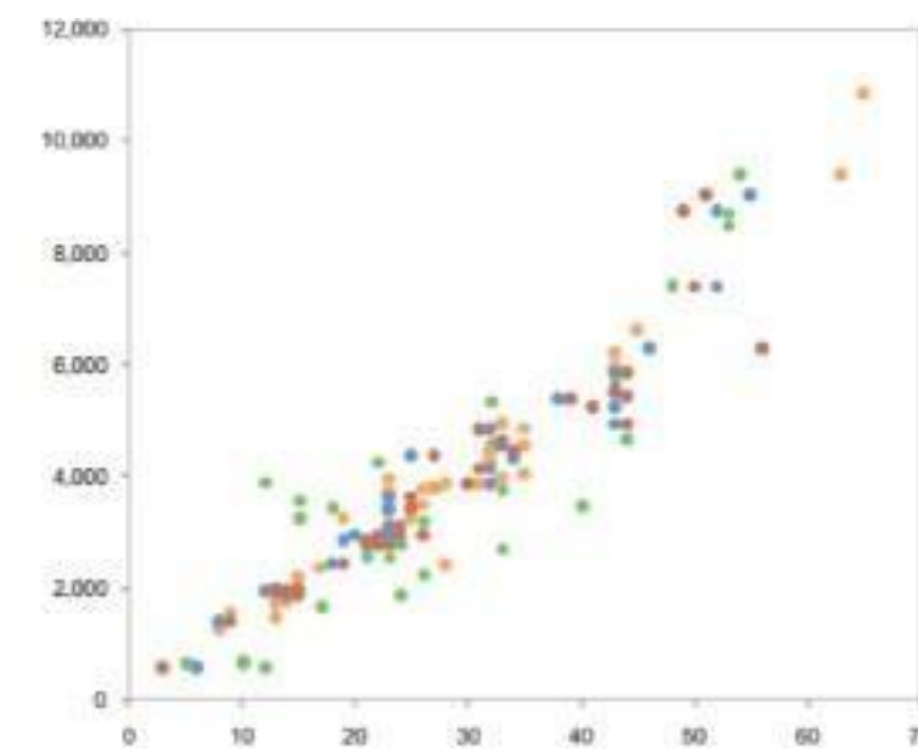
- Пастельные, ненасыщенные цвета лучше подходят для крупных элементов визуализации.
- Яркие, насыщенные - для мелких, в том числе, надписей, тонких линий.



✓ Цвета различимы



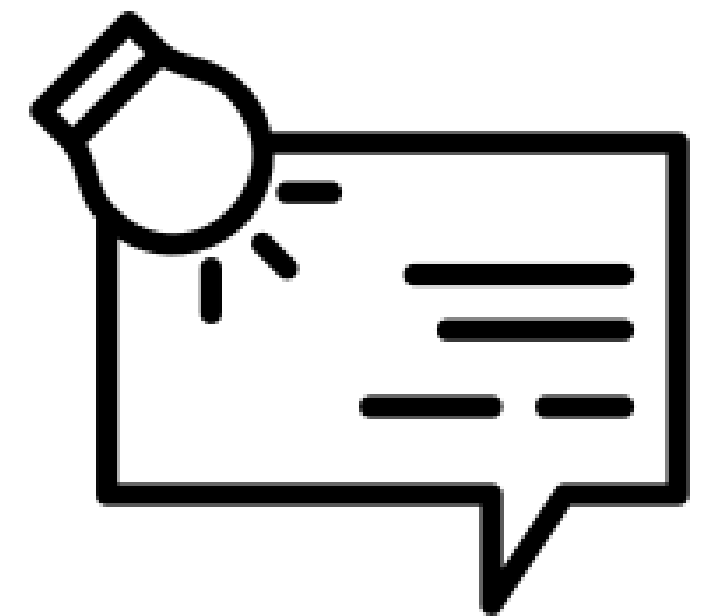
✗ Сложно различить элементы



✓ Увеличили масштаб и добавили насыщенность – стало лучше

Расставление акцентов

- При визуализации важны акценты в виде **направления, формы, размера, длины и ширины линии, цвета или насыщенности**
- Чтобы информация, которая имеет наибольшую значимость или релевантность (наиболее заметный стимул), отображалась для дальнейшего и более полного анализа путем сознательной обработки.



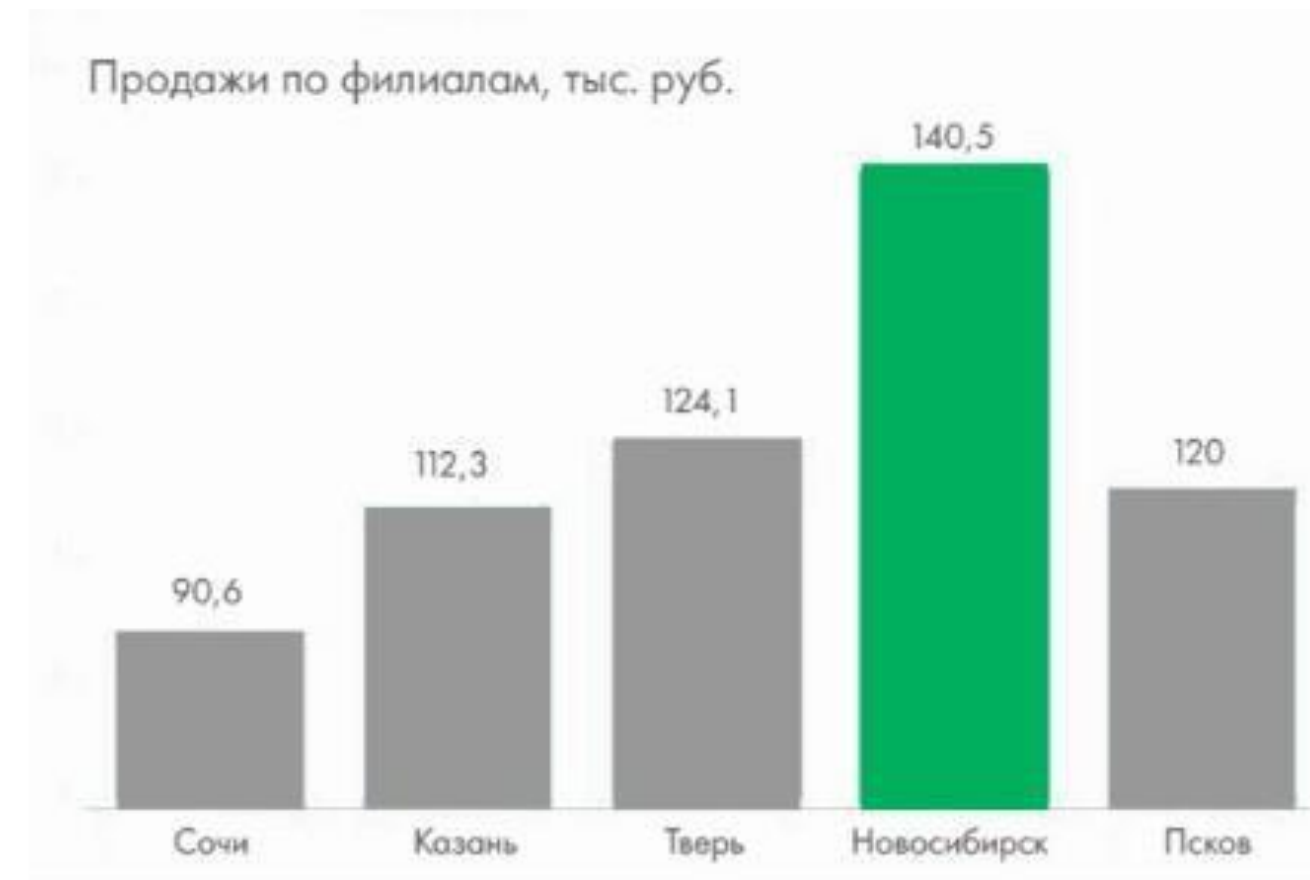
Визуальные атрибуты для расставления акцентов



Визуальные атрибуты для расставления акцентов

Длина и цвет

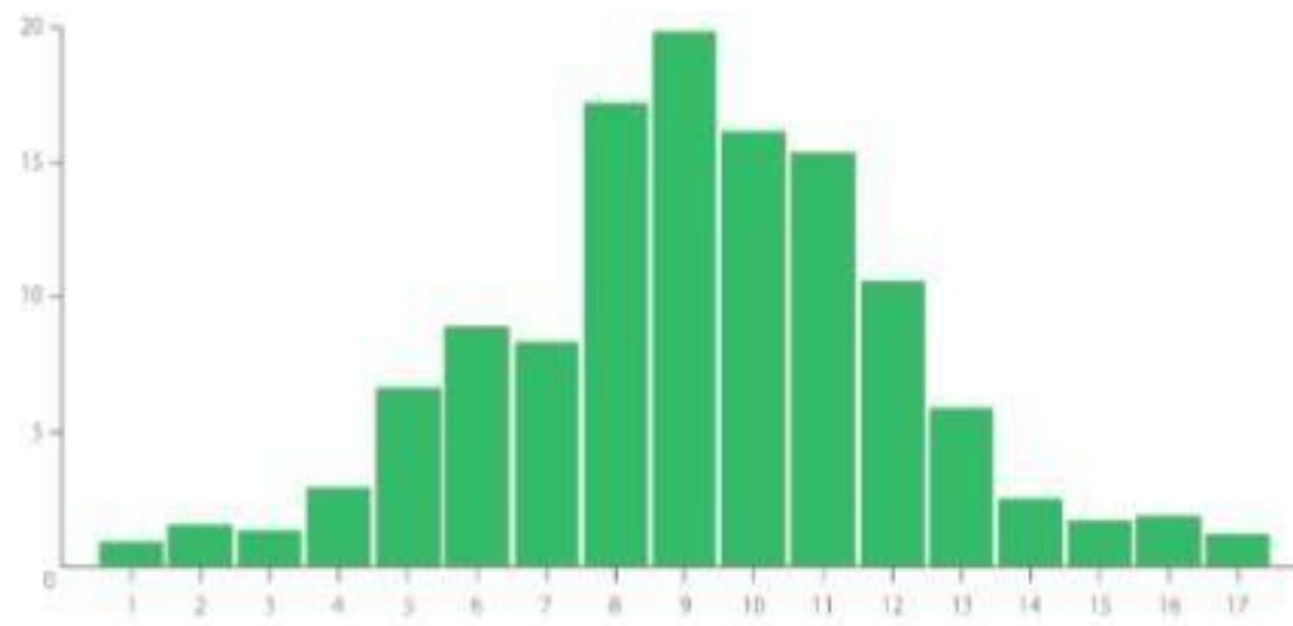
- Использование разной длины столбцов в диаграмме автоматически направляет наше внимание на город с большим объемом продаж.
- Этот эффект еще больше усиливается за счет использования цвета.



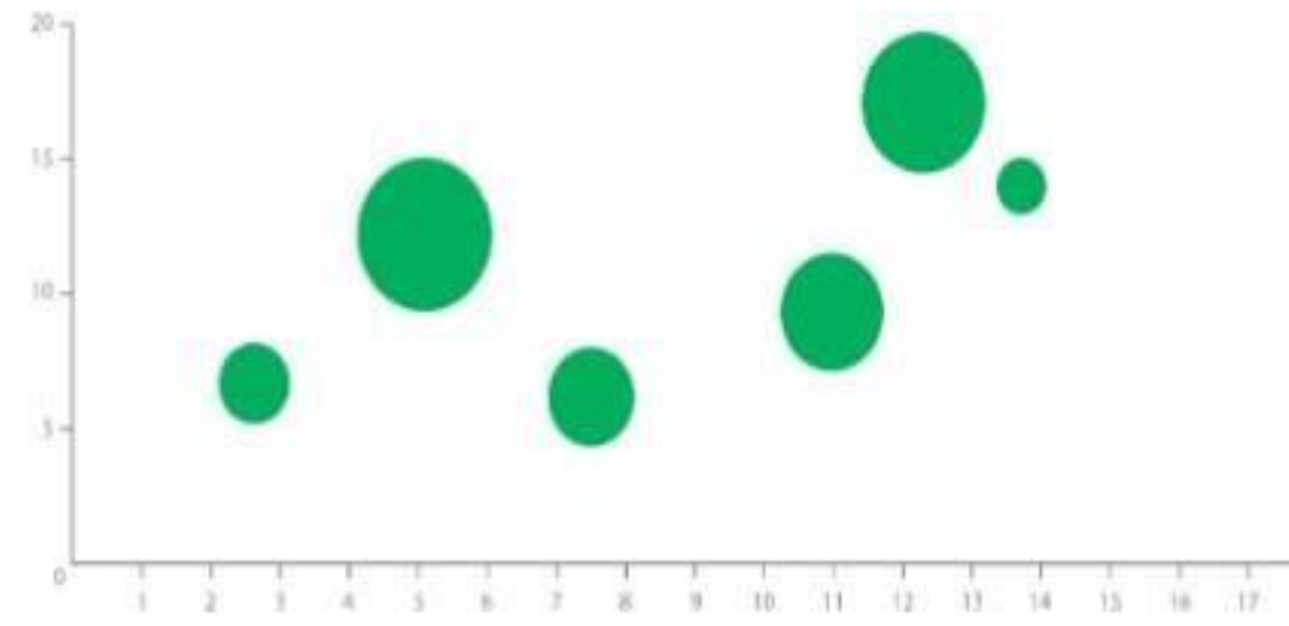
Визуальные атрибуты для расставления акцентов

Влияние на точность

- Стандартная гистограмма, где столбцы расположены на одной шкале, позволяет оценивать конкретные значения и делать более точные выводы
- А пузырьки разного размера или цвета на пузырьковой диаграмме дают лишь общие представления о разнице показателей.



Точные значения



Общее представление

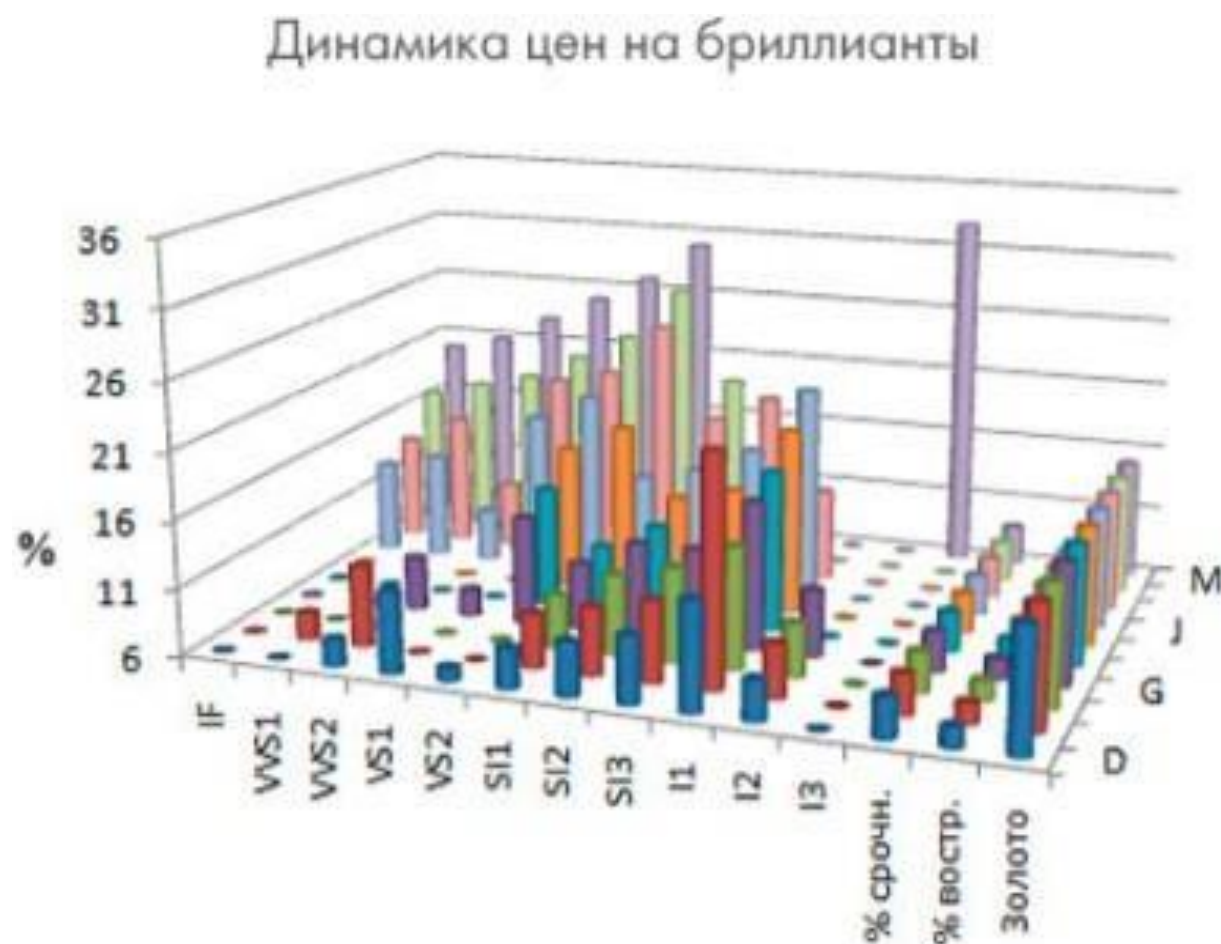
Не перегружать

- Визуализация должна помочь пользователю **сравнить два или более показателей** друг с другом.
- Дизайн визуализаций не должен препятствовать **пониманию связи** в данных или искажать данные.
- Излишнее количество элементов, цветов, попытки уместить в одну многоярусную диаграмму все имеющиеся данные приведет к тому, что визуальный ряд будет перегружен.



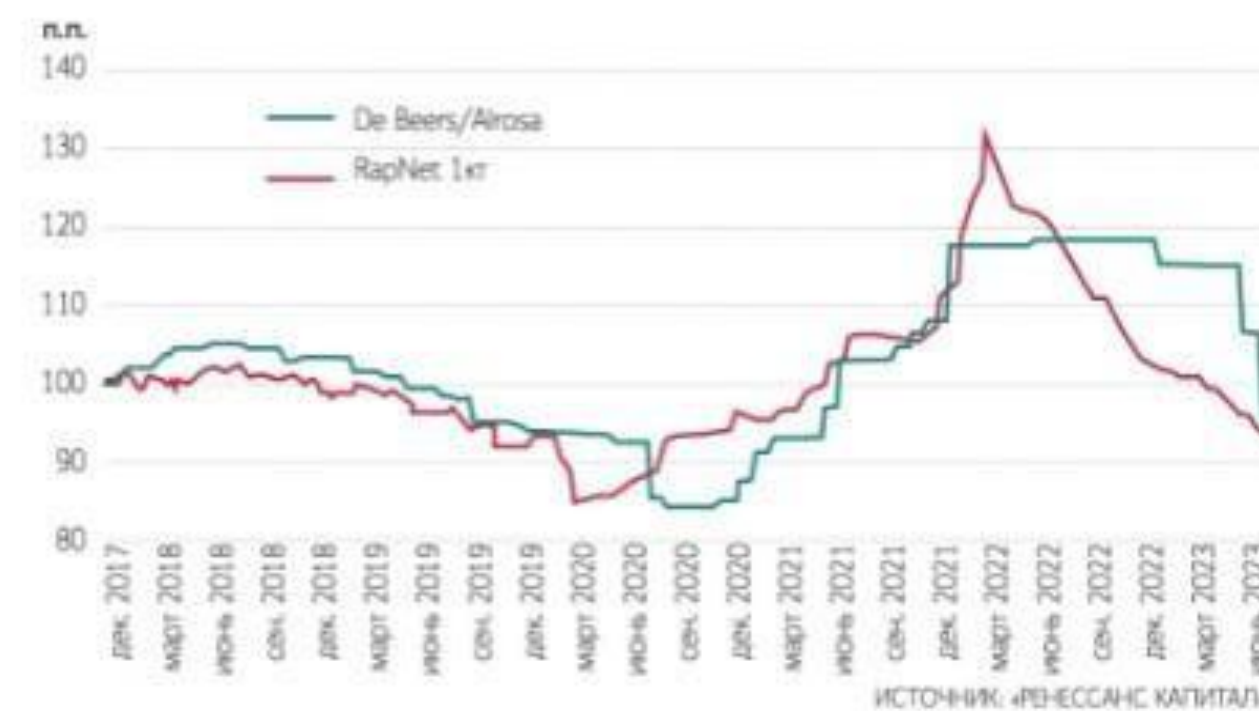
Не перегружать

- Если необходимо визуализировать сразу несколько типов данных и категорий, лучше разделить эту информацию на несколько диаграмм и графиков.



✗ Ничего не понятно

Как менялись индексы цен на бриллианты



✓ Интересно

Библиотеки визуализации данных Используемые в этом модуле

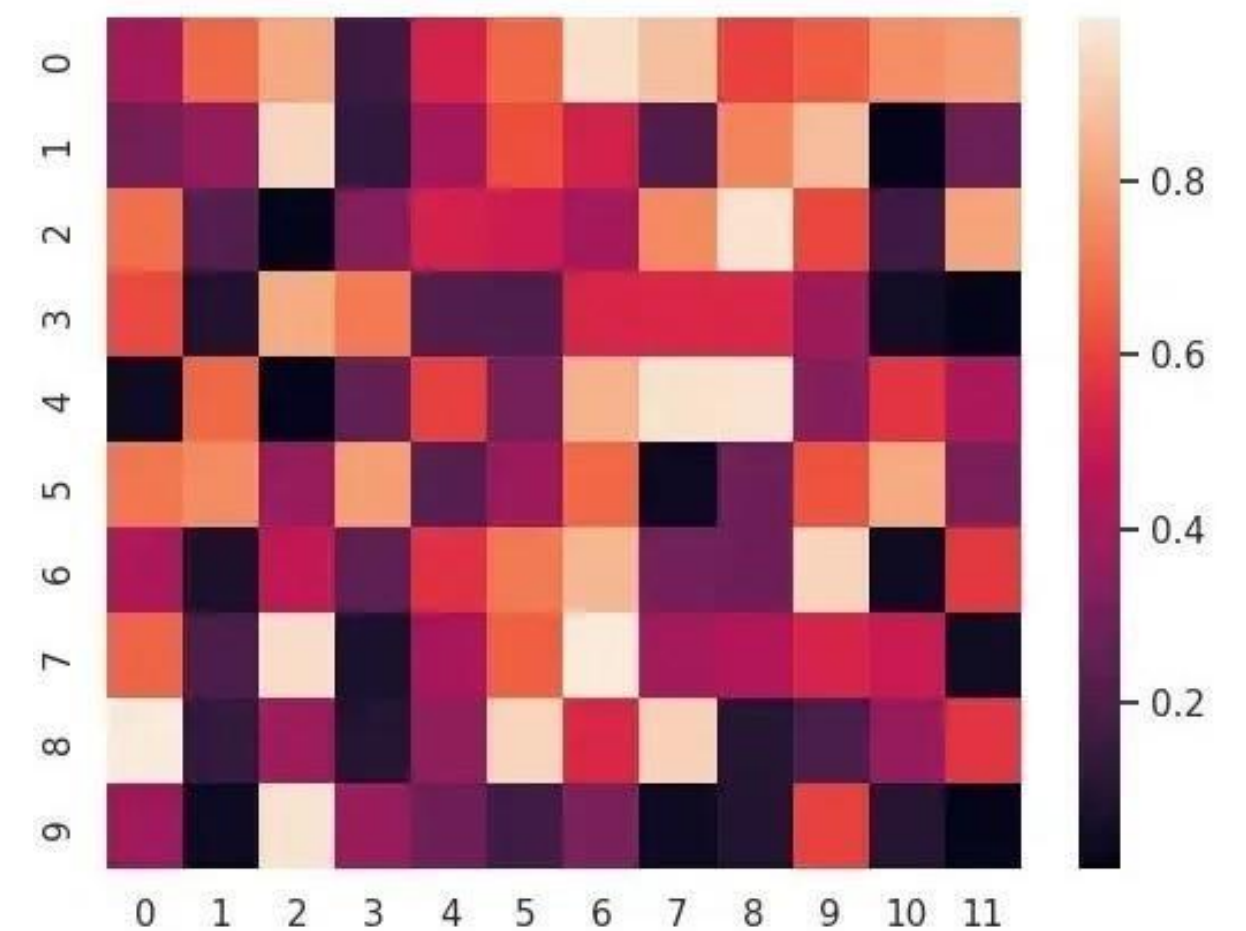
- Python (Matplotlib, Seaborn)

The logo for Matplotlib, featuring the word "matplotlib" in a blue sans-serif font. The "plot" part is replaced by a circular icon containing a stylized plot with several colored lines (orange, yellow, green, blue) and a black dot.The logo for Seaborn, featuring a circular icon on the left containing a stylized bar chart with blue bars and a light blue background. To the right of the icon, the word "seaborn" is written in a blue sans-serif font.

Типы Визуализации Данных

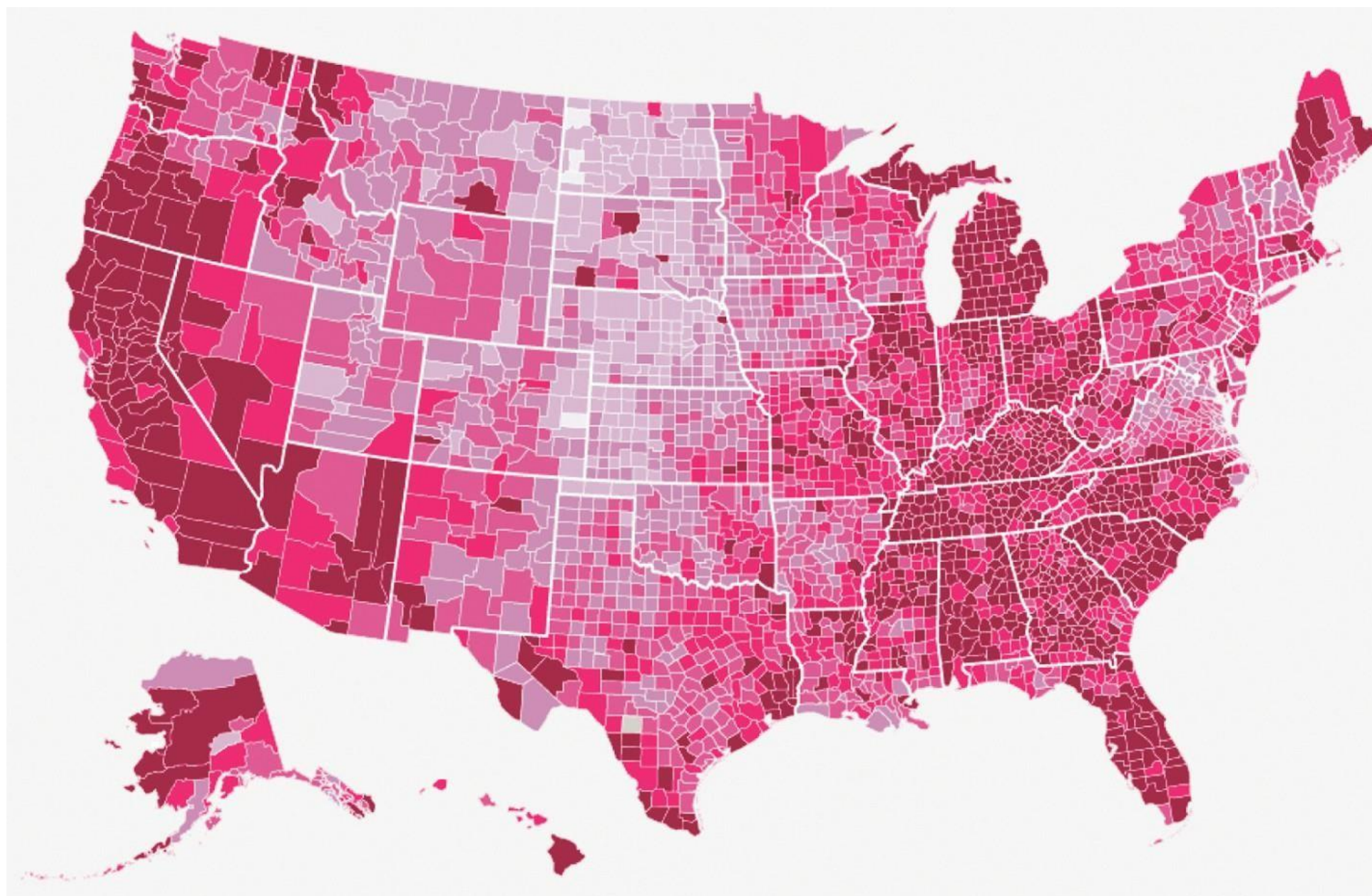
Тепловые карты (Heatmaps)

- Тепловая карта — это визуализация данных, которая отображается на карте или в таблице и **использует вариации разных цветов** для представления данных.
- Тепловая карта значений гораздо более удобна для просмотра и анализа **больших и сложных объемов данных с одного взгляда**.
- Полезна для **отображения плотности данных или насыщенности** значений в разных регионах или категориях.



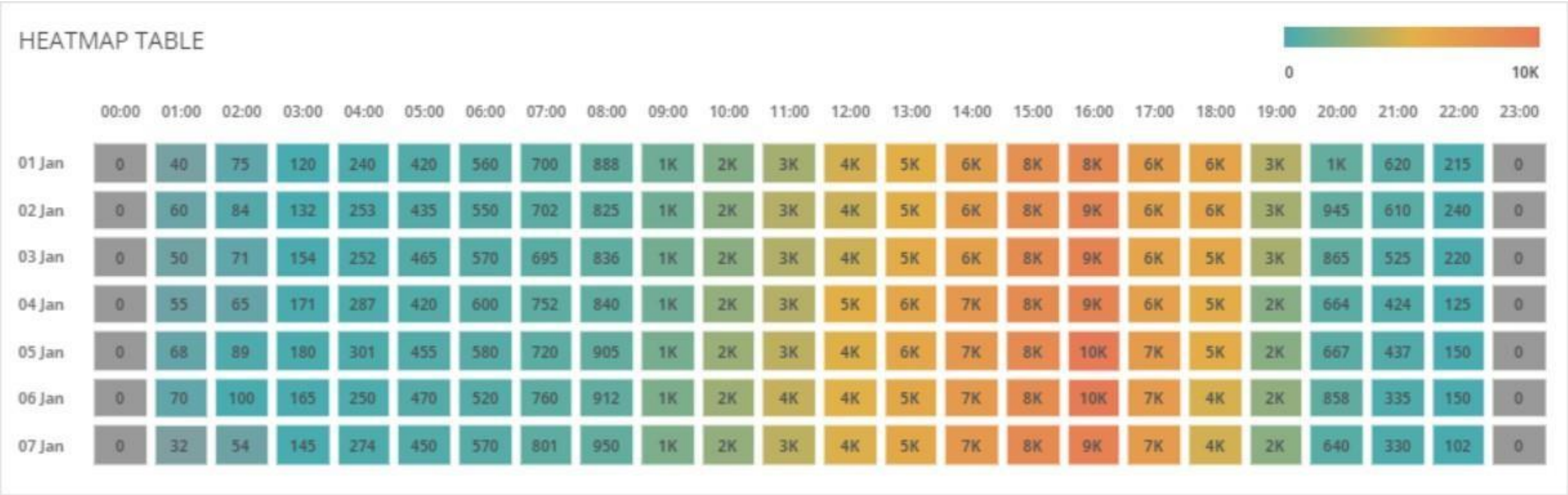
Тепловые карты

Пример



Тепловые карты

Пример



Визуализация тепловой карты в Python

```
# импортирование библиотек
import numpy as np
import seaborn as sn
import matplotlib.pyplot as plt

# генерация матрицы 10x10 чисел генерированных случайным образом
# от 1 до 100

data = np.random.randint(low = 1, high = 100, size = (10, 10))
print("Данные которые будут визуализированы:\n")
print(data)

# отрисовка тепловой карты
hm = sn.heatmap(data = data)

# показать тепловую карту
plt.show()
```


Визуализация тепловой карты в Python

Показать значения

```
# импортирование библиотек
import numpy as np
import seaborn as sn
import matplotlib.pyplot as plt

# генерация матрицы 10x10 чисел генерированных случайным образом
# от 1 до 100
data = np.random.randint(low=1,
                          high=100,
                          size=(10, 10))

# параметр для показа значений
annot = True

# отрисовка тепловой карты
hm = sn.heatmap(data=data,
                 annot=annot)

# показать тепловую карту
plt.show()
```

Визуализация тепловой карты в Python

Изменение палетки цветов

```
# импортирование библиотек
import numpy as np
import seaborn as sn
import matplotlib.pyplot as plt

# генерация матрицы 10x10 чисел генерированных случайным образом
# от 1 до 100
data = np.random.randint(low=1,
                          high=100,
                          size=(10, 10))

# задаем параметр для палетки цветов
cmap = "tab20"

# отрисовка тепловой карты
hm = sn.heatmap(data=data,
                 cmap=cmap)

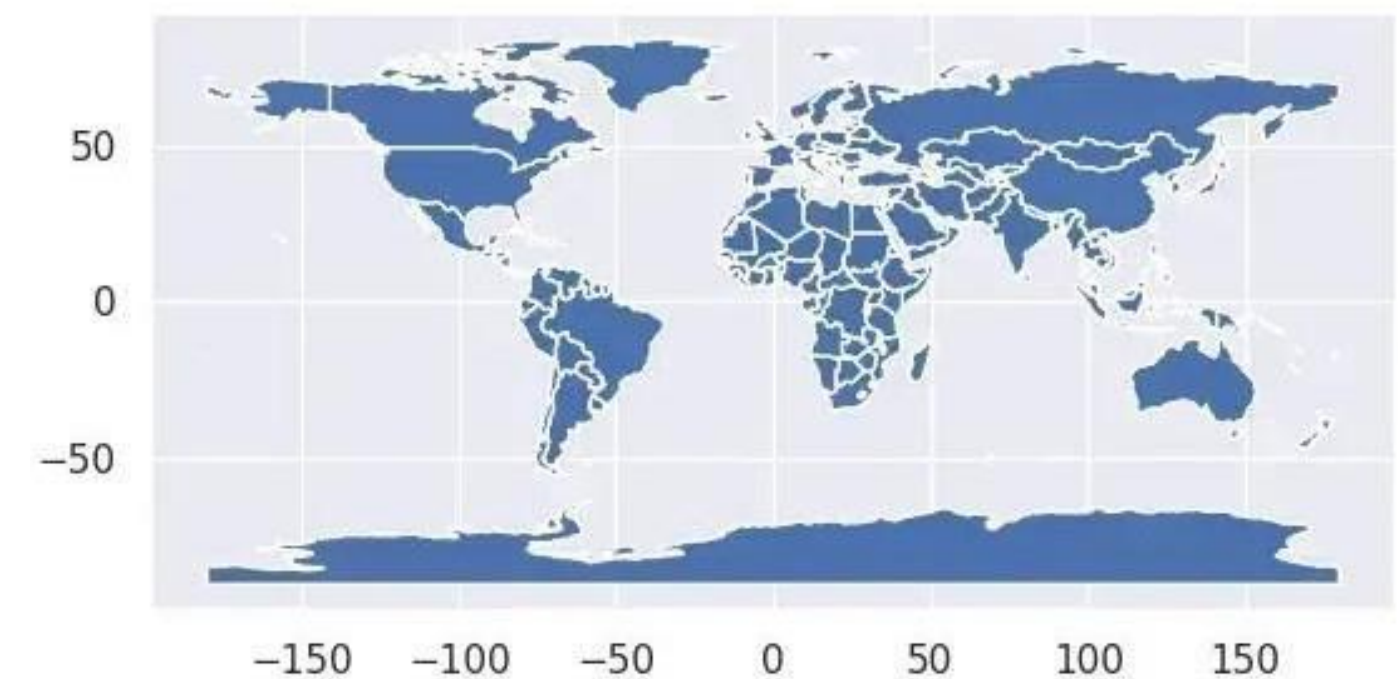
# показать тепловую карту
plt.show()
```

Показать все палетки цветов в Seaborn

```
import seaborn as sns  
  
# Напечатать наименования всех палеток цветов  
  
print(sns.palettes.SEABORN_PALETTE.keys())
```


Геопространственные карты

- Геопространственные карты **интегрируют географические данные с аналитическими данными**, обеспечивая пространственное измерение визуализации.
- Этот тип визуализации имеет решающее значение **для любых данных с географическим компонентом**, от глобальных экономических тенденций до планирования местных событий.
- Интеграция крупномасштабных географических данных с традиционными наборами данных **для обеспечения пространственного анализа**.



Установка библиотек в Python

Для работы с геопространственными данными

```
pip install folium
```

Визуализация геопространственной карты в Python

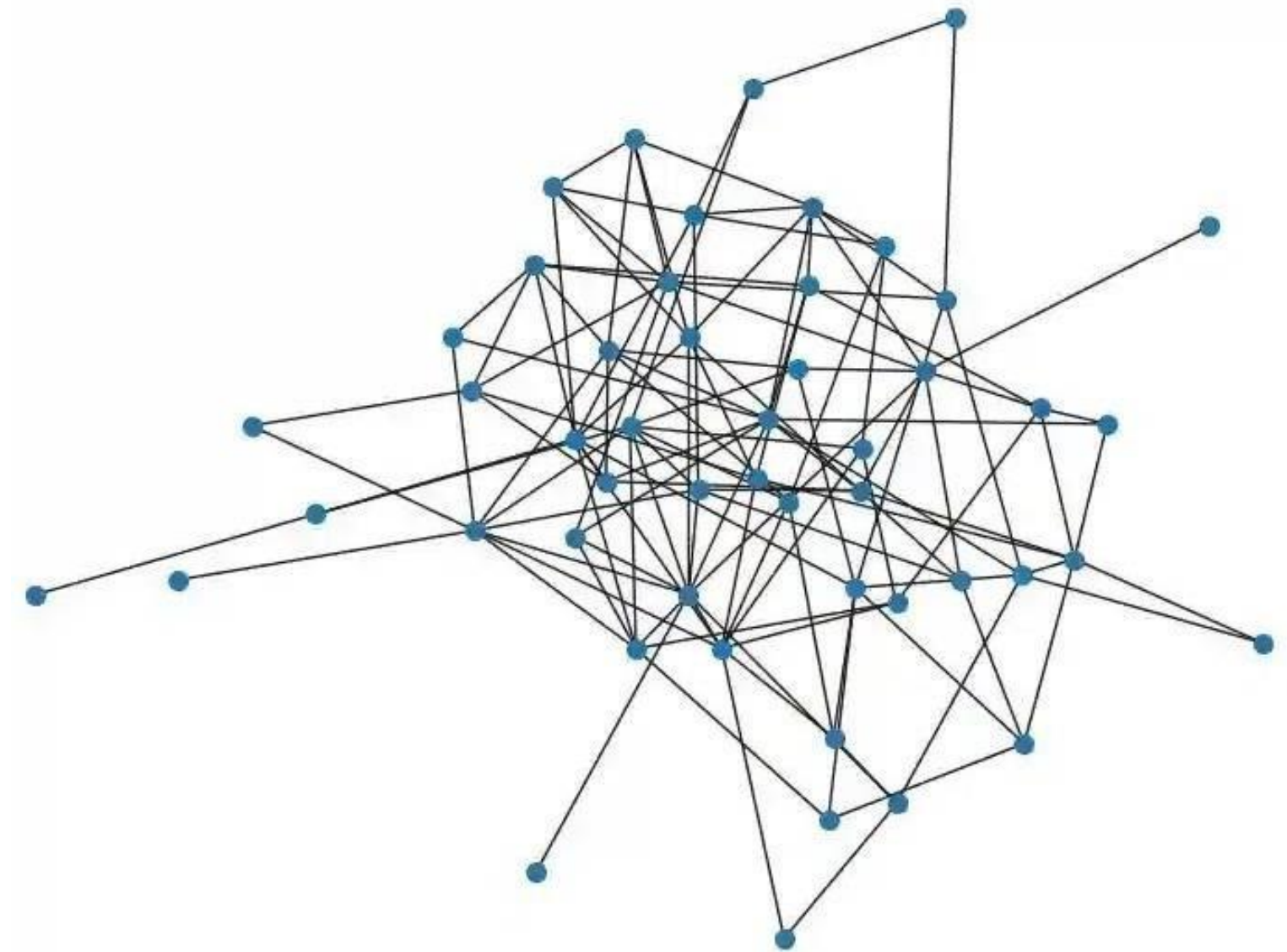
```
# импортирование библиотек folium, pandas
import folium
import pandas as pd

# инициализация карты
m = folium.Мар(location = [40, -95], zoom_start = 4)

# показать карту
m.save('my_map.html')
```


Сетевые диаграммы

- Сетевые диаграммы особенно ценны **для визуализации сложных взаимосвязей** между сущностями, что делает их незаменимыми в таких областях, как **анализ социальных сетей, инфраструктура интернета и биоинформатика**.
- Эти диаграммы состоят из **узлов** (представляющих сущности) и **ребер** (представляющих отношения).
- Идеально подходят для **визуализации отношений и взаимодействий** в данных, таких как социальные связи или потоки данных.



Применение визуализации данных

Здравоохранение

- **Пример:**

- Университет Джонса Хопкинса создал панель мониторинга COVID-19 с использованием визуализации больших данных для отслеживания глобального распространения COVID-19 в режиме реального времени.
- Эта панель отображала данные об уровнях инфицирования, смертности, показателях выздоровления и развертывании вакцинации, что позволяло медицинским работникам и политикам принимать обоснованные решения и эффективно реагировать на пандемию.

Применение визуализации данных

Финансы

- **Пример:**
 - JPMorgan Chase использует визуализацию больших данных для мониторинга рыночных тенденций и обнаружения мошеннических действий.
 - Их инструменты визуализации могут анализировать огромные объемы торговых данных в режиме реального времени, помогая выявлять необычные закономерности или транзакции, которые могут указывать на мошенничество.
 - Это позволяет банку принимать быстрые меры для снижения рисков и защиты активов своих клиентов.

Применение визуализации данных

Розничная торговля

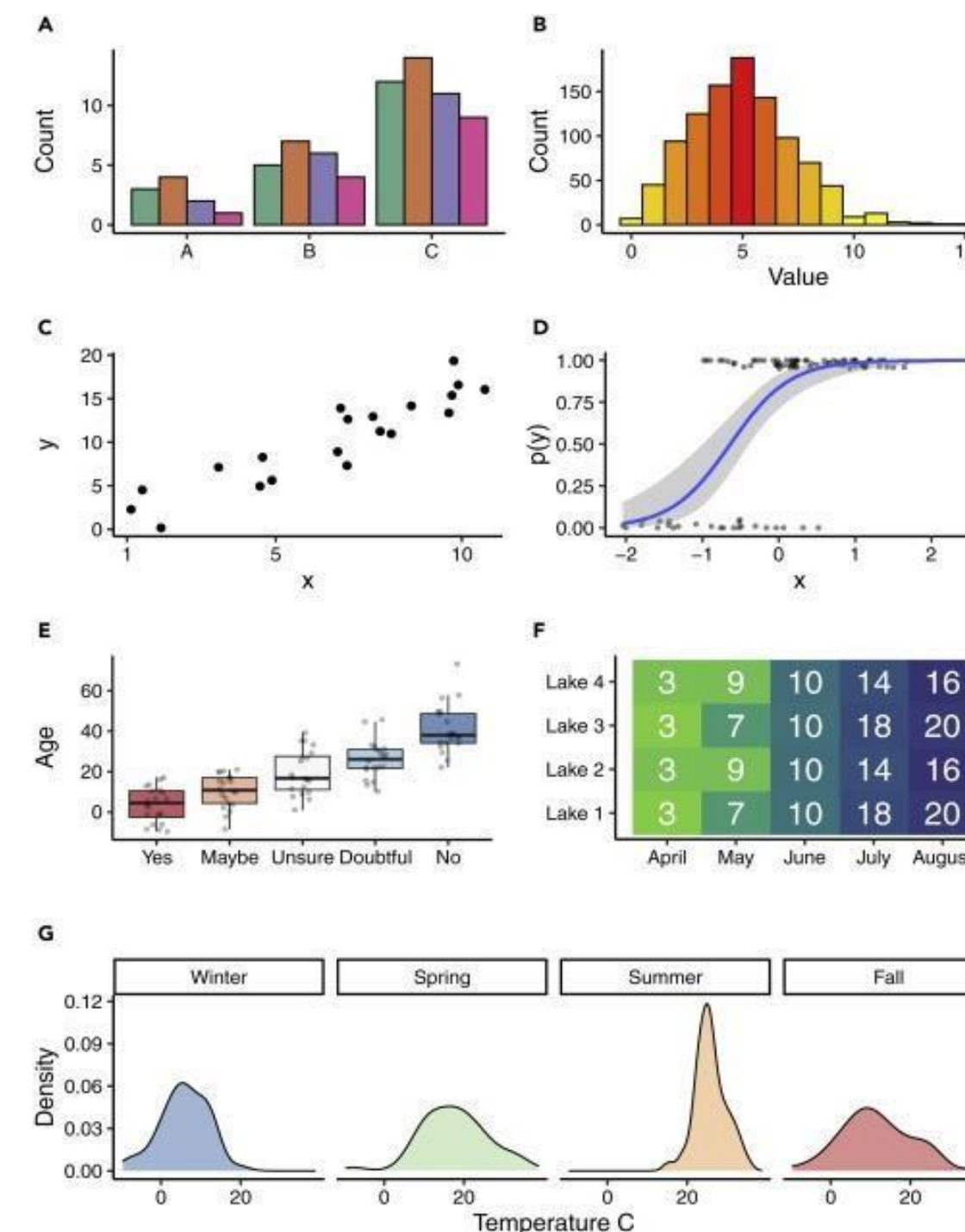
- **Пример:**

- Walmart использует визуализацию больших данных для оптимизации своей цепочки поставок и улучшения стратегий продаж.
- Анализируя данные о продажах, уровнях запасов и предпочтениях клиентов с помощью интерактивных панелей мониторинга, Walmart может определять тенденции, прогнозировать спрос и оптимизировать свою логистику, чтобы гарантировать доступность продуктов там и тогда, где и когда они нужны.

Применение визуализации данных

Как создать хорошие визуализации?

- **Выбрать диаграмму**, правильно отображающую конкретный тип данных
- **Использовать цвета** так, чтобы они помогли пользователю быстрее оценить ключевые показатели
- **Расставить акценты** в формах, размерах, с помощью насыщенности цветов, которые подчеркнут самое важное
- **Выбрать такой дизайн визуализаций**, который не будет препятствовать пониманию связи в данных или искажать их



Применение визуализации данных

Библиотеки визуализации данных

Используемые в этом модуле

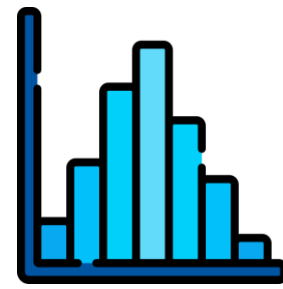
- Python (Matplotlib, Seaborn)



Применение визуализации данных

Пройденные типы визуализаций

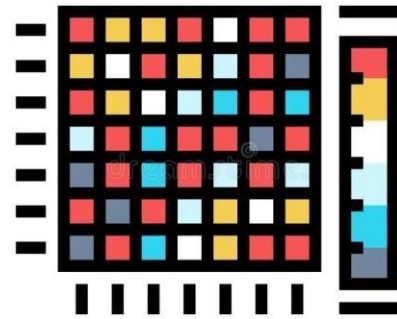
- Гистограмма



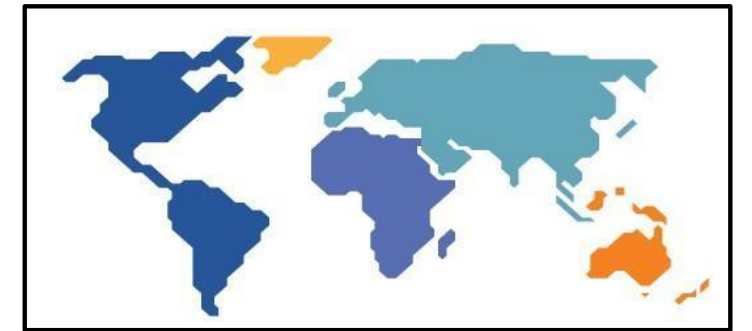
- Линейная диаграмма



- Тепловая карта



- Геопространственные карты



- Сетевые диаграммы



Применение визуализации данных

Этапы подготовки данных

- Разведочный Анализ Данных (Exploratory Data Analysis, EDA)

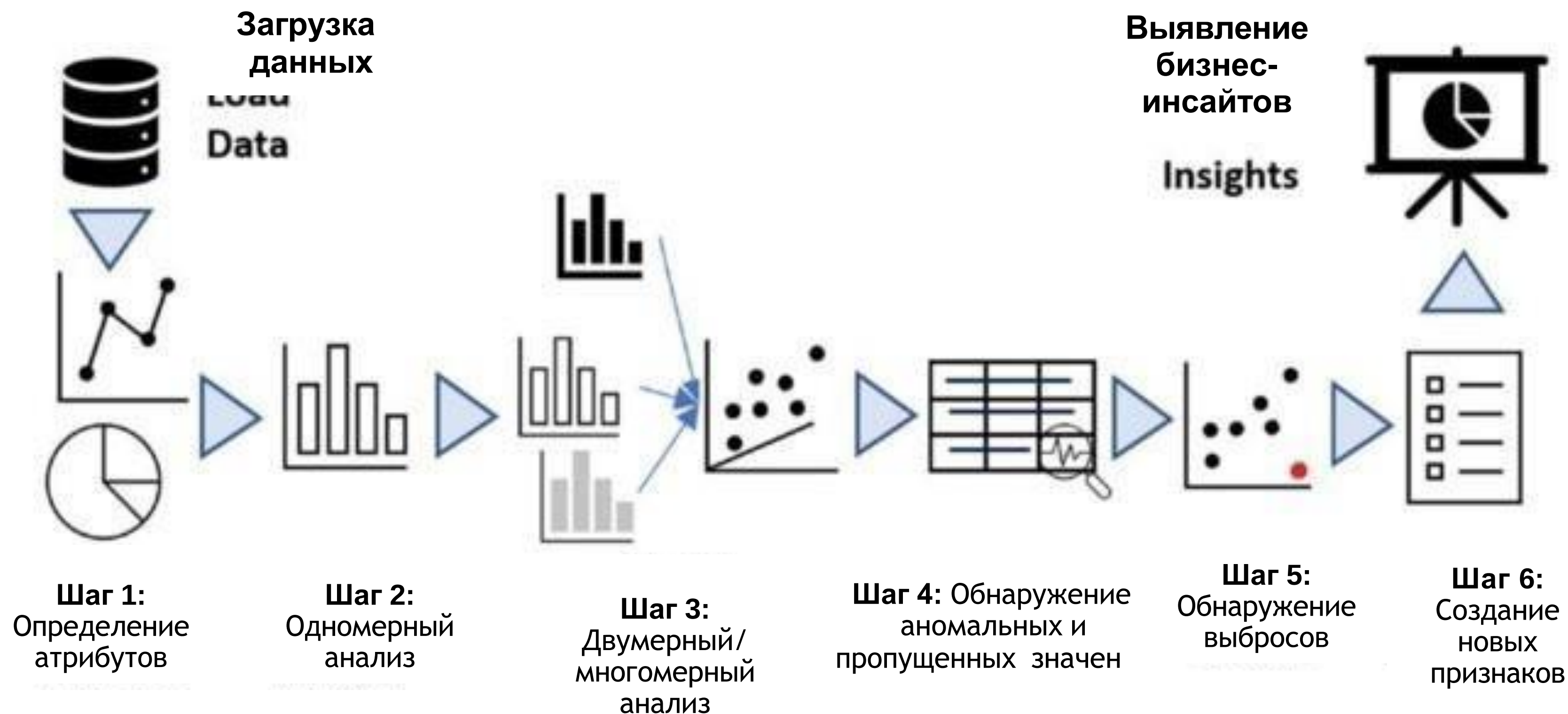


- Чистка данных



Применение визуализации данных

Разведочный Анализ данных



Применение визуализации данных

Разведочный анализ данных

Пример на population.csv

```
import pandas as pd
import matplotlib.pyplot as plt import seaborn
as sns

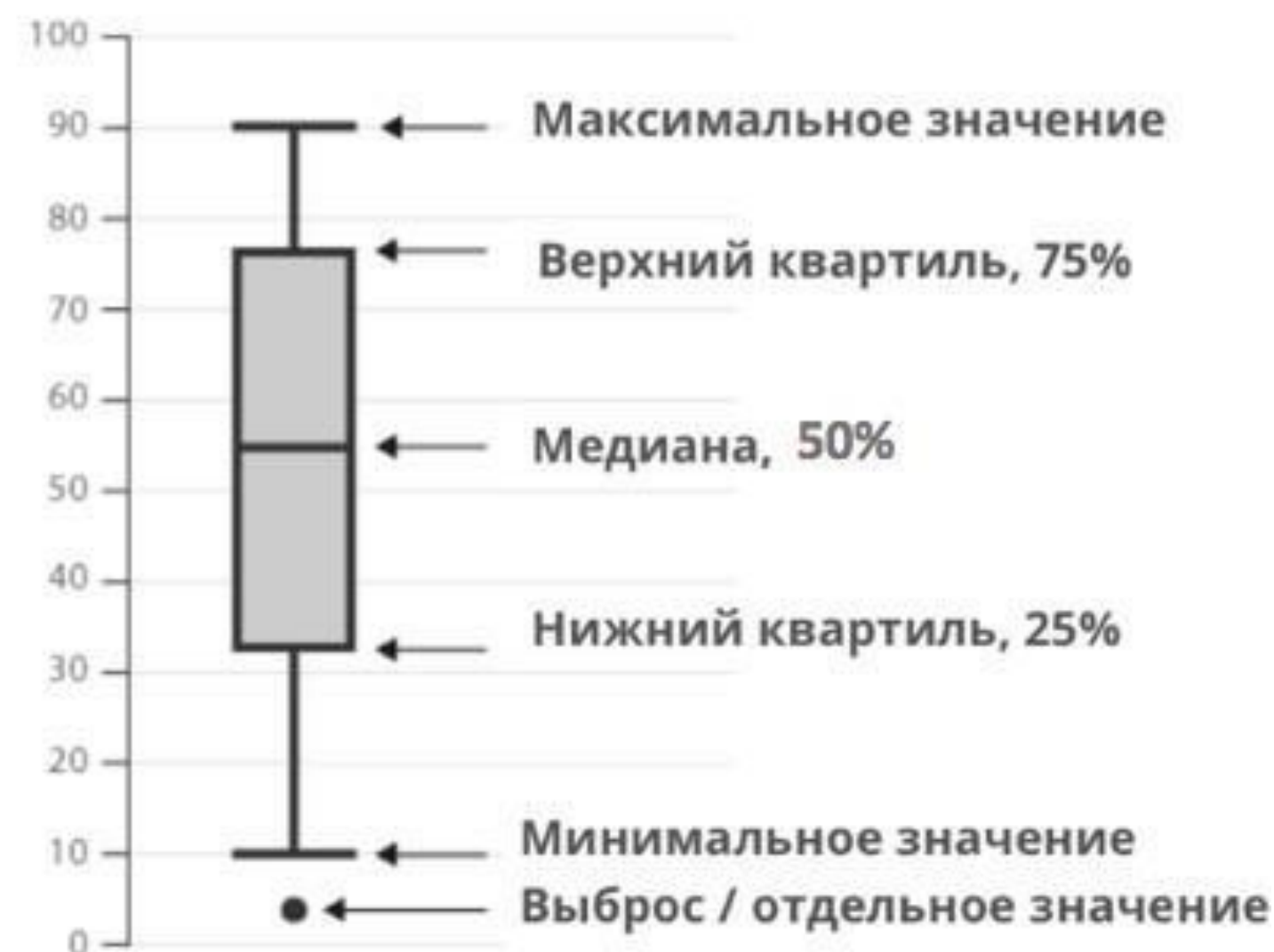
# 1. Гистограмма распределения Общего населения
plt.figure(figsize=(10, 5))
sns.histplot(df["Общее население"], bins=20, kde=True, color='skyblue') plt.xlabel("Общее
население")
plt.ylabel("Частота")
plt.title("Гистограмма распределения Общего населения") plt.show()
```

Применение визуализации данных

Ящичковая диаграмма/Бокс-плот

[Бокс-плот]

- * Показывает основную статистику выборки данных
- * Используется для быстрой оценки данных
- * Компактная и понятная



Применение визуализации данных

Ящичковая диаграмма/Бокс-плот

2. Ящичковая диаграмма - Распределение населения

```
plt.figure(figsize=(8, 5))
```

```
sns.boxplot(x=df["Общее население"], color="lightcoral") plt.xlabel("Общее население")
```

```
plt.title("Ящичковая диаграмма общего населения") plt.show()
```

Применение визуализации данных

Круговая диаграмма

3. Круговая диаграмма – Распределение по полу

```
gender_counts = [df["мужчины"].sum(),
df["женщины"].sum()] labels = ["Мужчины",
"Женщины"]

colors = ["blue", "pink"]

plt.figure(figsize=(6, 6))

plt.pie(gender_counts, labels=labels, autopct="%1.1f%%",
colors=colors, startangle=140)

plt.title("Распределение по полу в населении")

plt.show()
```

Применение визуализации данных

Столбчатая диаграмма

4. Столбчатая диаграмма - Сравнение городского и сельского населения

```
urban_population = df["городское население"].sum() rural_population =  
df["сельское население"].sum()
```

```
plt.figure(figsize=(8, 5))
```

```
sns.barplot(x=["Городское население", "Сельское население"], y=[urban_population,  
rural_population], palette=["green", "brown"])
```

```
plt.ylabel("Общее население")
```

```
plt.title("Сравнение городского и сельского населения") plt.show()
```

Применение визуализации данных

Точечная диаграмма

5. Точечная диаграмма - Мужчины vs Женщины

```
plt.figure(figsize=(7, 5))
```

```
sns.scatterplot(x=df["мужчины"], y=df["женщины"], alpha=0.7, color="purple")
```

```
plt.xlabel("Мужчины")
```

```
plt.ylabel("Женщины")
```

```
plt.title("Точечная диаграмма: Мужчины vs Женщины") plt.show()
```


Спасибо за внимание!