

# Статистический Анализ и Визуализация Данных

## Тема: ПО и современные инструменты для статистического анализа данных и создания визуализаций

Student: Kuzmitkaia Maia  
Grupa: SD-232

Profesor: conf. univ. dr. Galina Marusic

## План:

Введение;

Программы для работы с таблицами;

Сравнение Excel и Google Sheets;

Языки программирования;

Сравнение R и Python;

BI-системы;

Сравнение Tableau и Power BI;

Выводы;

Библиография

## Введение

В современном мире данных становится все больше, и их грамотный анализ играет ключевую роль в принятии обоснованных решений.

Без качественного анализа данные остаются просто числами, но благодаря современным инструментам их можно превратить в ценные инсайты, которые помогают бизнесу, науке и государственному управлению принимать эффективные решения. Графики, дашборды и интерактивные отчеты позволяют быстро оценивать ситуацию, находить проблемы и разрабатывать стратегии их решения.

Использование мощных инструментов, таких как **Excel**, **Google Sheets**, **Python**, **R**, **Tableau** и **Power BI**, делает процесс анализа данных более удобным, точным и автоматизированным. Выбор правильного инструмента зависит от задач, объема данных и требований к аналитике.

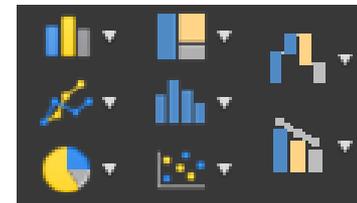
# Программы для работы с таблицами

## Excel

- это один из самых популярных инструментов для работы с данными. Он широко используется в бизнесе, финансах, науке и других сферах благодаря своей доступности и мощным возможностям анализа.

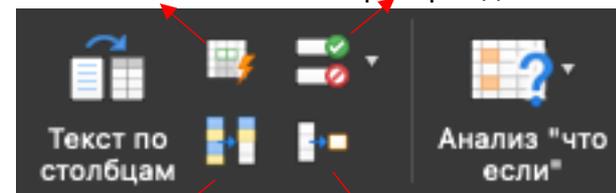
## Возможности

- ✓ Сводные таблицы
- ✓ Функции и формулы
- ✓ Графики и диаграммы
- ✓ Power Query
- ✓ Макросы и VBA
- ✓ Анализ «Что Если?»



Мгновенное заполнение

Проверка данных



Удаление дубликатов

Консолидация

# Программы для работы с таблицами

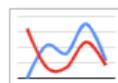
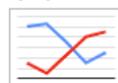
## Google Sheets

- это облачный сервис для работы с таблицами, который предлагает мощные возможности анализа данных и удобные инструменты для совместной работы. Он особенно популярен среди команд, которым важно редактировать и анализировать данные в режиме реального времени.

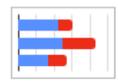
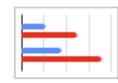
### Возможности

- Сводные таблицы
- Функции и формулы
- Графики и диаграммы
- Google Apps Script
- Совместная работа

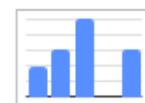
Графики



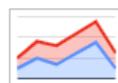
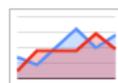
Линейчатые



Другие



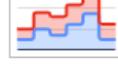
С областями



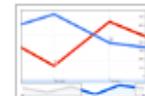
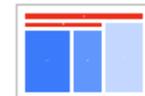
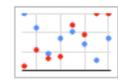
Круговые



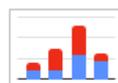
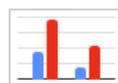
Точечные



Точечные



Столбчатые



Карта



A	B	C	D
14	25	36	47
25	38	47	58
38	47	58	60

## Сравнение Excel и Google Sheets

Excel	Критерий	Google Sheets
Устанавливается на ПК, требует лицензию	Доступность	Работает в браузере, бесплатен
Ограниченная, через OneDrive	Совместная работа	Позволяет совместное редактирование
Поддерживает большие файлы	Объем данных	Ограничен размером таблицы
Продвинутые функции анализа данных, макросы	Функциональность	Меньше функций, но есть скрипты Apps Script
VBA, Power Query	Автоматизация	Apps Script
Мощные надстройки и расширения	Гибкость	Хорош для базовых и средних задач

# Языки программирования



— это язык программирования, специально созданный для статистического анализа и визуализации данных. Он широко используется в науке, экономике, биостатистике и других областях, где требуется сложная обработка данных.

## Пакеты для анализа и визуализации

ggplot2 – создание графиков

tidyr – преобразование данных

readr – чтение и запись данных

dplyr – манипуляция данными

tibble – работа с тибблами

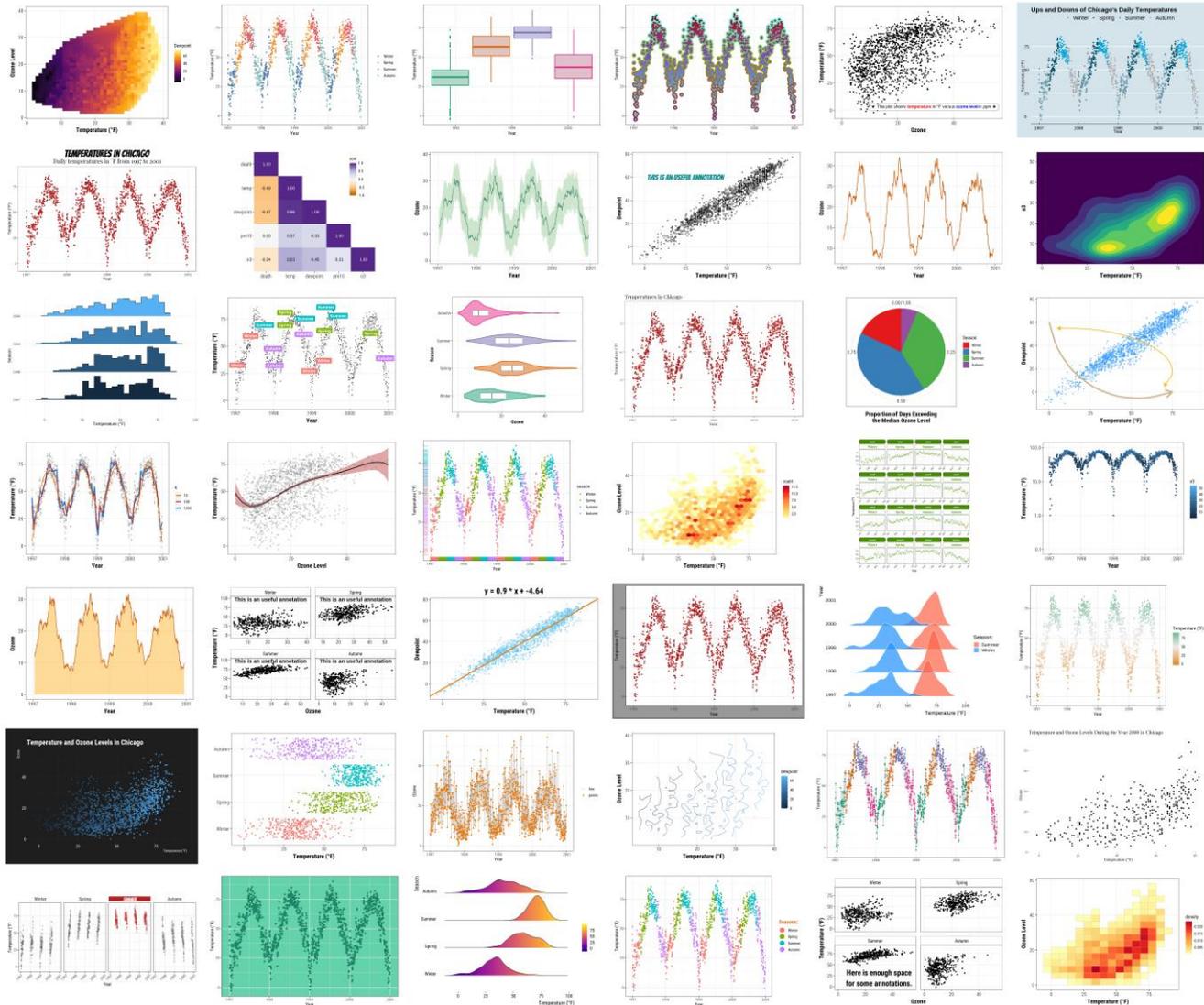
purrr – работа с функциями

stringr – работы со строками

forcats – работа с факторами



tidyverse





# Языки программирования

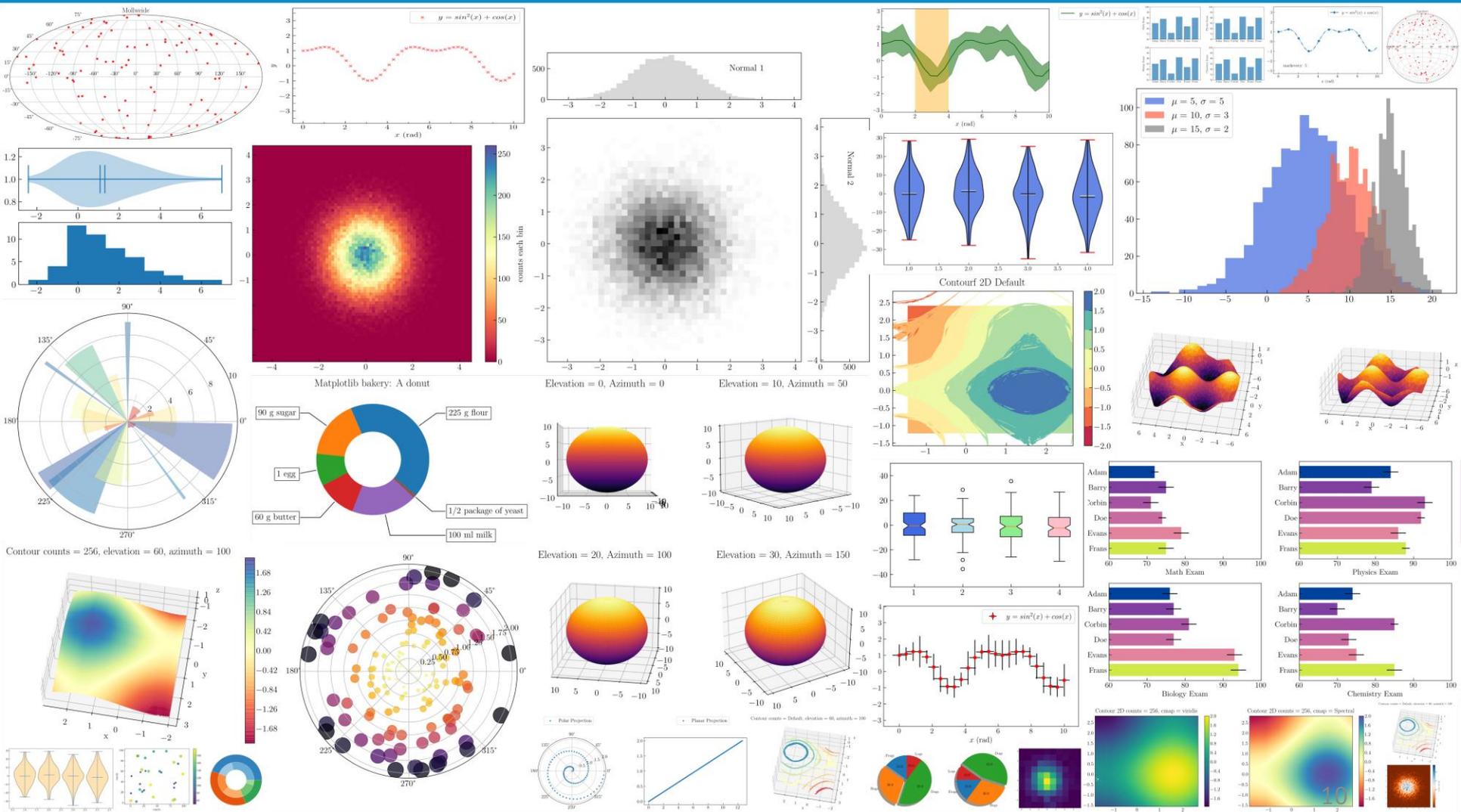
## Python

— это один из самых популярных языков программирования для анализа и визуализации данных. Он предлагает множество мощных библиотек, которые охватывают широкий спектр задач, от обработки данных до построения сложных визуализаций.

## Библиотеки для анализа и визуализации

**Анализ:** Pandas, NumPy, SciPy

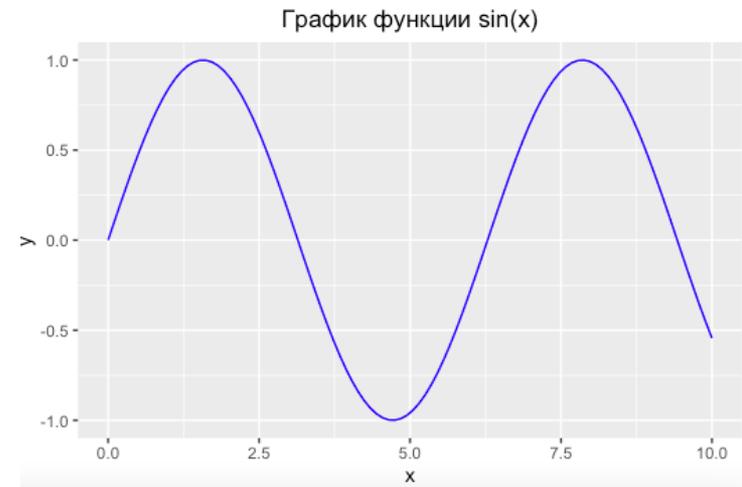
**Визуализация:** Matplotlib, Seaborn, Plotly



# Сравнение R и Python

R: `x <- seq(0, 10, length.out = 100)`  
`y <- sin(x)`

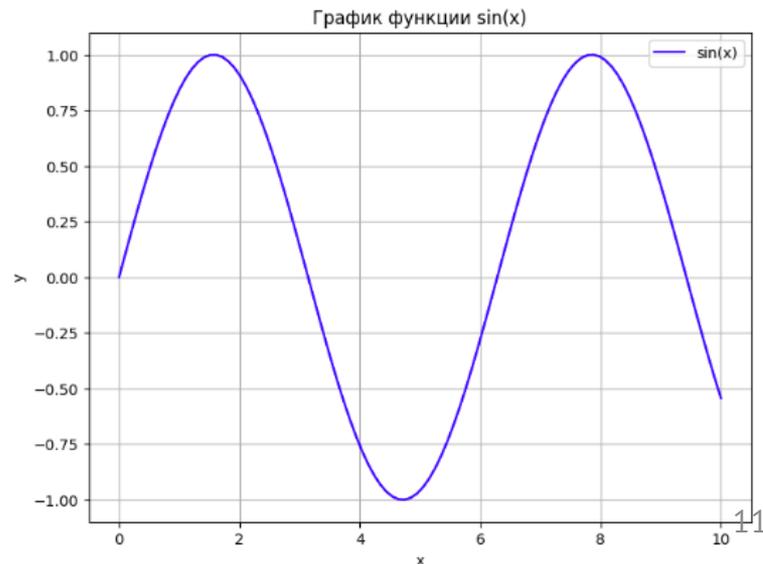
```
library(ggplot2)
data <- data.frame(x, y)
ggplot(data, aes(x = x, y = y)) +
  geom_line(color = 'blue') +
  labs(title = "График функции sin(x)", x = "x", y = "y") +
  theme(plot.title = element_text(hjust = 0.5))
```



Python: `import numpy as np`  
`import matplotlib.pyplot as plt`

```
x = np.linspace(0, 10, 100)
y = np.sin(x)
```

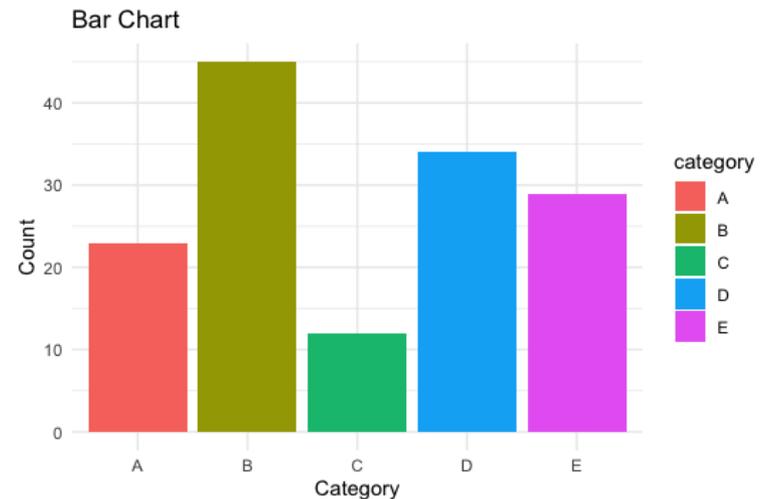
```
plt.figure(figsize=(8, 6))
plt.plot(x, y, label='sin(x)', color='blue')
plt.title('График функции sin(x)')
plt.xlabel('x')
plt.ylabel('y')
plt.legend()
plt.grid(True)
plt.show()
```



# Сравнение R и Python

R:

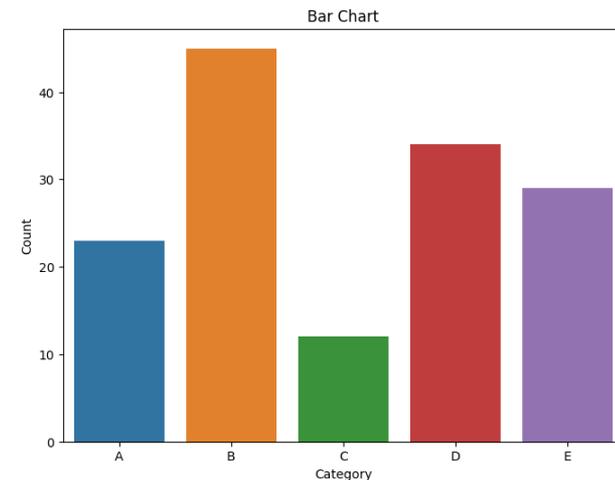
```
library(ggplot2)
df <- data.frame(
  category = c("A", "B", "C", "D", "E"),
  count = c(23, 45, 12, 34, 29)
)
ggplot(df, aes(x = category, y = count, fill = category)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Bar Chart in R", x = "Category", y = "Count")
```



Python:

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

df = pd.DataFrame({
  'category': ['A', 'B', 'C', 'D', 'E'],
  'count': [23, 45, 12, 34, 29]
})
plt.figure(figsize=(8, 6))
sns.barplot(x='category', y='count', data=df,
  palette="tab10")
plt.title("Bar Chart in Python")
plt.xlabel("Category")
plt.ylabel("Count")
plt.show()
```



## Сравнение R и Python

R	Критерий	Python
Статистика, визуализация, научные исследования	Основное применение	Анализ данных, машинное обучение, автоматизация
Специализированный синтаксис, требует обучения	Простота освоения	Достаточно просто, если знать основы
Tidyverse	Манипуляция, визуализация	Pandas, Matplotlib, Seaborn
Есть ограничения	Большие данные	Хорошая поддержка
RStudio	IDE	Jupyter Notebooks, Spyder
Популярен в академических кругах	Популярность	Используется в IT, науке, финансах



## BI-системы

- это программное обеспечение, предназначенное для сбора, анализа, обработки и визуализации данных с целью поддержки принятия управленческих решений.

## Tableau

- это один из ведущих инструментов бизнес-аналитики, используемый для визуализации данных. Он позволяет пользователям легко создавать интерактивные дашборды и отчеты без необходимости глубоких знаний программирования.







## Варианты графиков

**Data** Analytics

Basic Table\_data

Search

**Tables**

- Country
- Measure Names
- Happiness Score
- Year of Year
- Basic Table\_data.csv (Co...
- Latitude (generated)
- Longitude (generated)
- Measure Values

**Columns** Longitude (generated)

**Rows** Latitude (generated)

**Filters** Country

**Marks** Automatic

Color Size Label

Detail Tooltip

AVG(Happines..)

Country

AVG(Happines..)

### Europe Happiness

Country	Happiness Score
Denmark	7.500
Norway	7.433
Finland	7.333
Iceland	7.100
Switzerland	7.000
Austria	6.900
Germany	6.933
France	6.500
Spain	6.500
Italy	5.867
UK	5.867
Poland	5.867
Czechia	5.733
Slovenia	5.500
Lithuania	5.500
Latvia	5.500
Estonia	5.500
Belgium	5.500
Netherlands	5.500
Portugal	5.400
Greece	5.400
Cyprus	5.400
Malta	5.400
Croatia	5.100
Slovakia	5.100
Hungary	5.100
Slovenia	5.100
Bulgaria	5.100
Romania	5.100
Algeria	5.133
Libya	5.133
Egypt	5.133
Saudi Arabia	5.133
Iran	5.133
Pakistan	5.133
India	5.133
Kazakhstan	5.133
China	5.133
USA	5.133
Canada	5.133
Australia	5.133
New Zealand	5.133
South Africa	5.133
Brazil	5.133
Mexico	5.133
Argentina	5.133
Colombia	5.133
Venezuela	5.133
Peru	5.133
Chile	5.133
Ecuador	5.133
Bolivia	5.133
Paraguay	5.133
Uruguay	5.133
Costa Rica	5.133
Panama	5.133
Dominican Republic	5.133
Honduras	5.133
Nicaragua	5.133
Guatemala	5.133
Belize	5.133
Jamaica	5.133
Cuba	5.133
Haiti	5.133
Dominican Republic	5.133
Trinidad and Tobago	5.133
Suriname	5.133
Guayana Francesa	5.133
Greenland	5.133

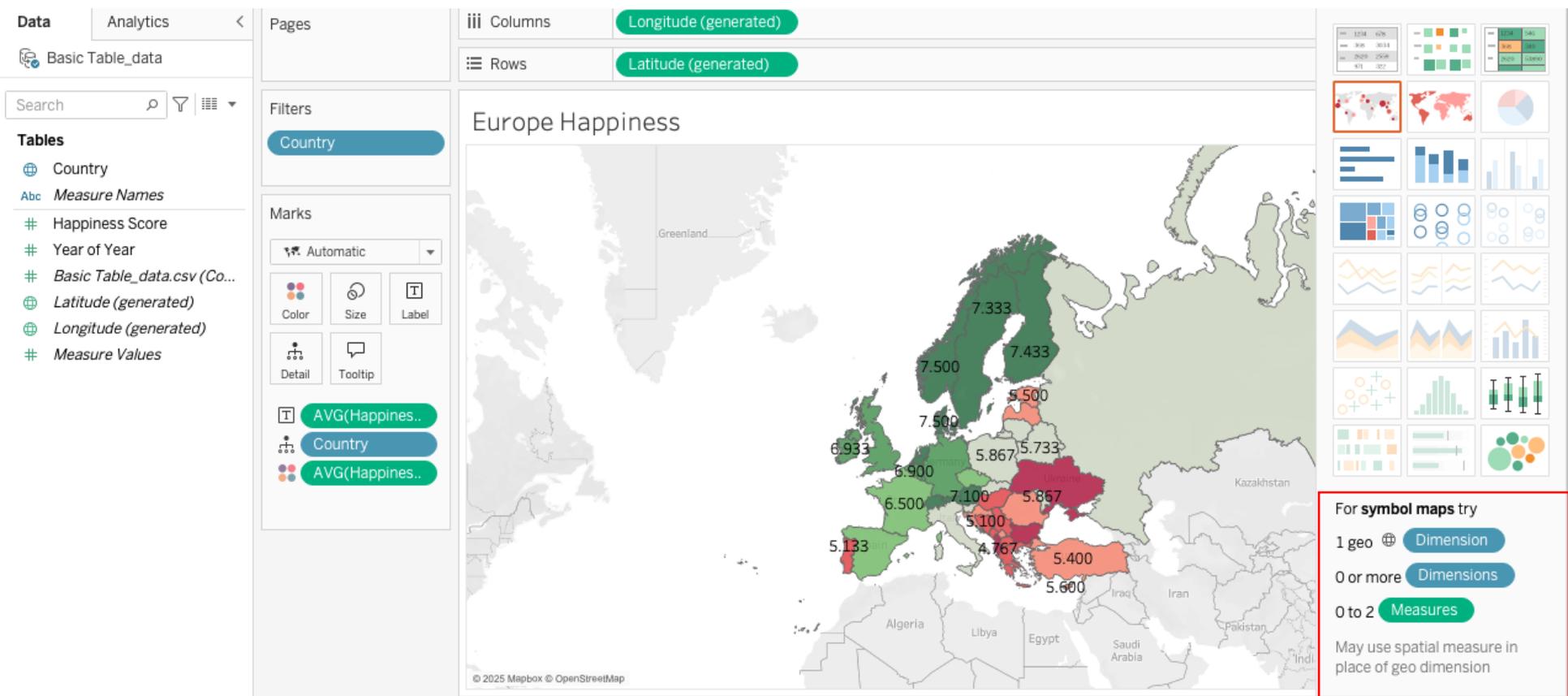
**For symbol maps try**

1 geo Dimension

0 or more Dimensions

0 to 2 Measures

May use spatial measure in place of geo dimension



**Data** Analytics

Basic Table\_data

Search

**Tables**

- Country
- Measure Names
- Happiness Score
- Year of Year
- Basic Table\_data.csv (Co...
- Latitude (generated)
- Longitude (generated)
- Measure Values

**Pages**

**Filters**

Country

**Marks**

Automatic

Color Size Label

Detail Tooltip

AVG(Happines..

Country

AVG(Happines..

**Columns** Longitude (generated)

**Rows** Latitude (generated)

### Europe Happiness Viz

Country	Score
Norway	7,500
Sweden	7,433
Denmark	7,333
Finland	7,100
Iceland	6,933
Ireland	6,900
Portugal	6,500
Spain	6,500
France	5,867
Germany	5,867
Poland	5,867
Czechia	5,733
Slovenia	5,500
Italy	5,500
Greece	5,500
UK	5,133
Belgium	5,100
Netherlands	5,100
Austria	5,100
Switzerland	5,100
Spain	5,100
Malta	4,767
Cyprus	4,767
Turkey	5,400
Romania	5,600

**Visualizations**

For **symbol maps** try

- 1 geo **Dimension**
- 0 or more **Dimensions**
- 0 to 2 **Measures**

May use spatial measure in place of geo dimension



## BI-системы

### Power BI

— это инструмент бизнес-аналитики (BI) от Microsoft, предназначенный для сбора, анализа и визуализации данных. Он интегрируется с другими продуктами Microsoft, такими как Excel, Azure и SQL Server, что делает его удобным для пользователей экосистемы Microsoft.

Clipboard: Paste, Cut, Copy, Format painter

Data: Get data, Excel workbook, Data hub, SQL Server, Enter data, Datasource, Recent sources

Queries: Transform data, Refresh

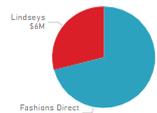
Insert: New visual, Text box, More visuals

Calculations: New measure, Quick measure

Sensitivity: Sensitivity, Publish

### Store Sales Overview

This Year Sales by Chain



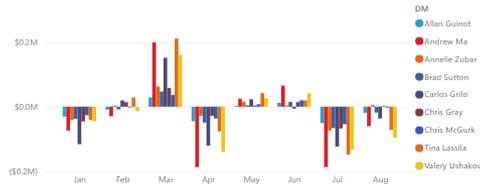
10  
New Stores

104  
Total Stores

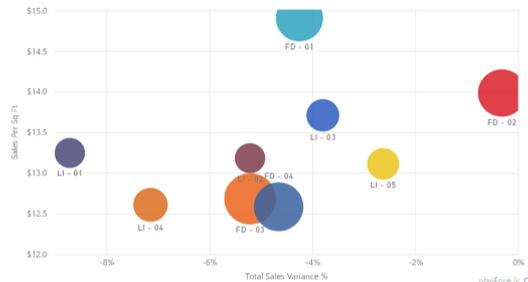
This Year Sales by PostalCode and Store Type



Total Sales Variance by FiscalMonth and District Manager



Total Sales Variance %, Sales Per Sq Ft and This Year Sales by District and District



### Filters

Search

#### Filters on this page

- Chain is (All)
- City is (All)
- District is (All)
- Name is (All)
- Open Month is (All)
- Store Type is (All)

### Visualizations

#### Build visual

Visual icons for various chart types: Bar, Line, Pie, Map, etc.

#### Values

Add data fields here

#### Drill through

Cross-report

Keep all filters

Add drill-through fields here

### Fields

Search

- Sales
- District
- Item
- Store
- Time

## Сравнение Tableau и Power BI

Tableau	Критерий	Power BI
Гибкий интерфейс, но требует обучения	Простота использования	Интуитивно понятный для пользователей Excel
Высокая, сложные визуализации	Интерактивность	Отлично работает с Microsoft инструментами
Дорого, подписка или лицензия (или Public version)	Стоимость	Более доступно, есть бесплатная версия
API, SQL, облачные сервисы	Интеграция	Полная интеграция с Microsoft (Excel, Azure)
Быстрее работает с большими объемами данных	Производительность	Хорошо оптимизирован для ограниченных объемов
Windows, MacOS	Платформа	Windows



## Выводы

### Таблицы

Хранение данных,  
маленькие задачи,  
быстрый анализ,  
простые визуализации

### R и Python

Большие данные,  
продвинутый анализ,  
статистические тесты,  
разнообразие  
визуализаций

### BI-системы

Интерактивные  
дашборды, продвинутая  
визуализация без кода,  
обоснование решений

## Библиография

1. Coursera Staff, Google Sheets vs. Excel: What's the Difference?, [дата обращения 02.02.2025], доступно <https://www.coursera.org/articles/google-sheets-vs-excel>
2. Coursera Staff, Python or R for Data Analysis: Which Should I Learn?, [дата обращения 02.02.2025] доступно <https://www.coursera.org/articles/python-or-r-for-data-analysis>
3. DataFlair Team, R Vs Python – The most trending debate of aspiring Data Scientists, [дата обращения 02.02.2025], доступно <https://data-flair.training/blogs/r-vs-python/>
4. Javier Canales Luna, Python vs R for Data Science: Which Should You Learn?, [дата обращения 02.02.2025], доступно <https://www.datacamp.com/blog/python-vs-r-for-data-science-whats-the-difference>
5. Shaun Edmond, Matt Crabtree, Power BI vs. Tableau: Which is The Better Business Intelligence Tool in 2024?, [дата обращения 02.02.2025], доступно <https://www.datacamp.com/blog/power-bi-vs-tableau-which-one-should-you-choose>
6. Zev Ross, Beautiful plotting in R: A ggplot2 cheatsheet, , [дата обращения 02.02.2025], доступно <http://zevross.com/blog/2014/08/04/beautiful-plotting-in-r-a-ggplot2-cheatsheet-3/>