

Python初級數據分析員證書

(五) 進階Python數據分析及可視化技巧

9. 簡介數據可視化工具

Visualization Tools

9. 簡介及比較數據可視化工具

Chapter Summary

- Introduction
- Matplotlib
- Seaborn
- Plotly
- Superset
- Tableau
- PowerBI



Introduction

Why visualization ?

What to visualize ?

Who is the audience ?

How to visualize ?

Type of communication

1. Verbal Communication

- Storytelling
- Conversation

2. Written Communication

- Paper Document
- Electronic Document
- Social Media

3. Physical Communication

- Body posture, eye contact, facial expression
- Emotion, tone, intonation, pitch

4. Visual Communication

- Graphs, illustration, drawing, sign
- Electronic resources

Impact on words vs visual

Unemployment rate sees eight month decline to 3.5 pc

Local | 19 Jan 2023 6:37 pm



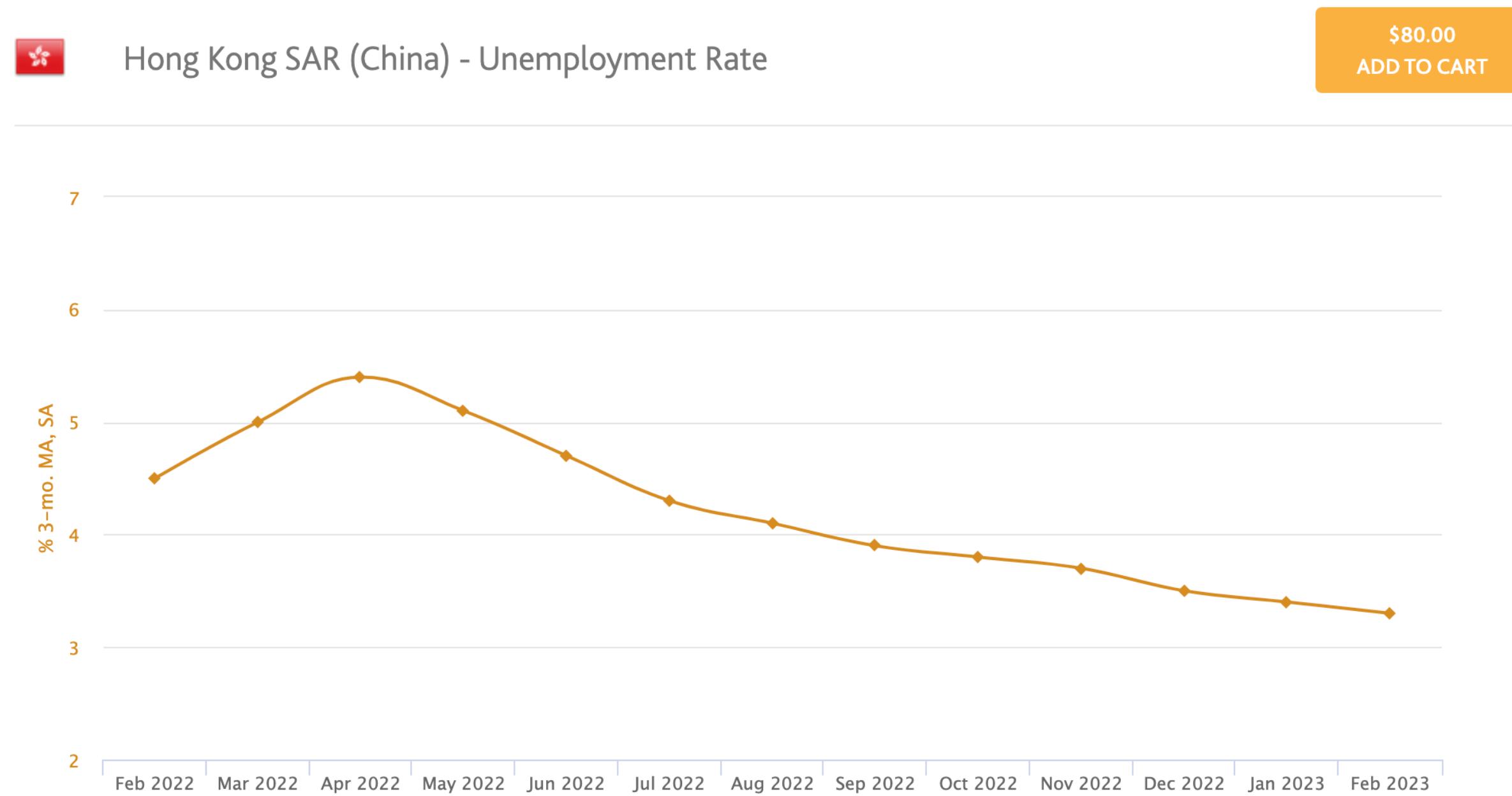
Hong Kong's unemployment rate logged a continuous declining trend for eight months to hit 3.5 percent, while around 126,000 workers are looking for jobs.

The Census and Statistics Department released the latest labor force statistics report on Thursday, showing the seasonally adjusted unemployment rate of the city decreased from 3.7 percent in the period from September to November last year to 3.5 percent between November and December.

The number of unemployed was also reduced by 12,700 to 126,000 people when compared to the last calculation period. The unemployment rate dropped across almost all the major economic sectors, with more distinct decreases observed in the construction, retail, transportation, and education sectors.

"The local labor market will see a further improvement with the cancellation of most Covid restrictions and the resumption of travels with the mainland," said Secretary for Labour and Welfare Chris Sun Yuk-han.

Impact on words vs visual



Hong Kong SAR (China): Unemployment Rate		
Mnemonic	LBR.IHKG	
Unit	% 3-mo. MA, SA	
Adjustments	Seasonally Adjusted	
	Monthly	
Data	Feb 2023	3.3
	Jan 2023	3.4

Series Information		
Source	Census and Statistics Department Hong Kong Special Administrative Region (HKSARG)	
Release	Employment Situation - Monthly	
Frequency	Monthly	
Start Date	10/31/1981	
End Date	2/28/2023	

Impact on words vs visual



Chairman's Statement

Following the strong economic rebound in the second half of 2020, global recovery continued during 2021 helped by the accelerated vaccination rollout programs and easing of lock-downs and economic reopenings in many parts of the world. However, the global recovery was not a smooth journey in 2021 as new COVID-19 variants continued to emerge during the year. Different policy responses in various markets led to operational and cost challenges exacerbated by pandemic-related supply-demand mismatches. Nevertheless, the Group was able to respond nimbly to changing economic and business environments in various sectors and geographies and is able to report pleasing results for the full year.

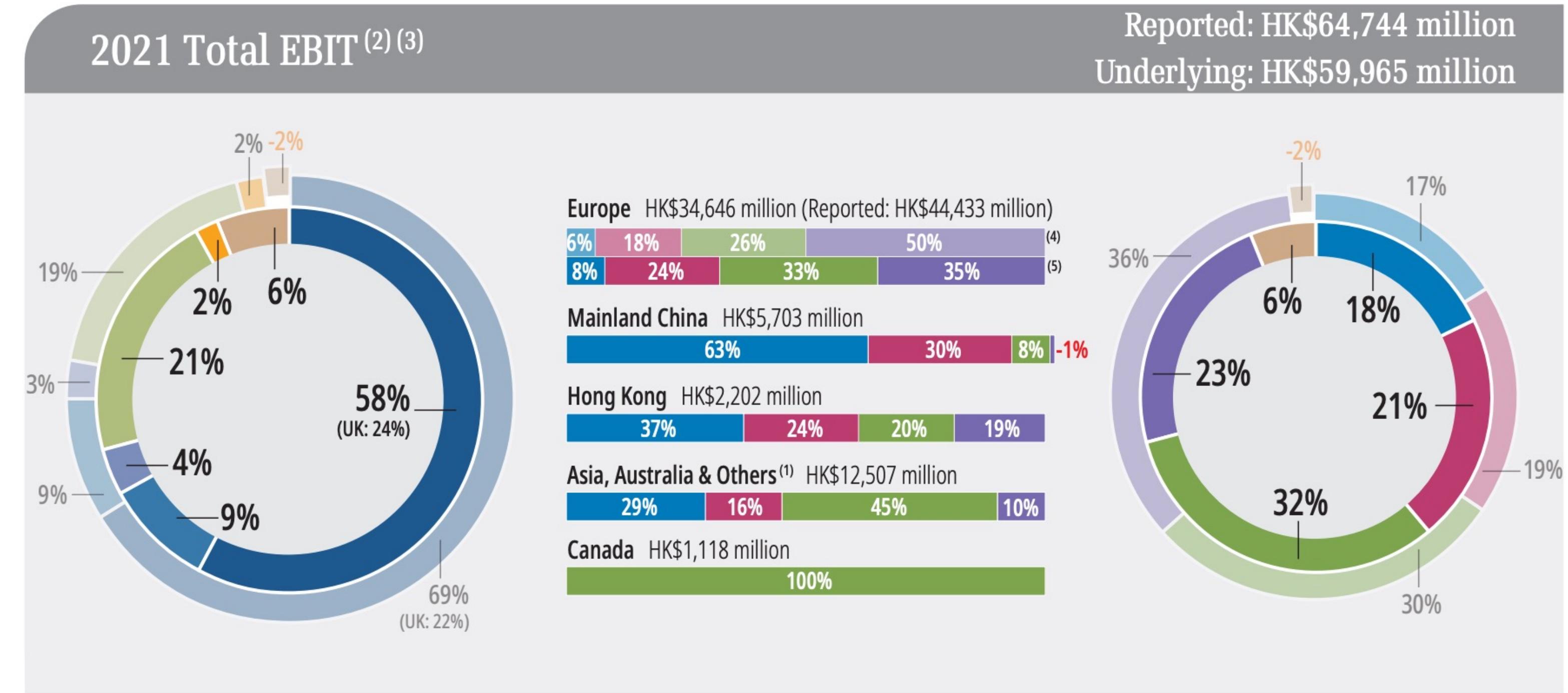
The Group reported EBITDA and EBIT growth of 15% and 20% respectively in reported currency compared to last year was primarily driven by improvements in the Ports and Retail divisions together with positive underlying results of Cenovus Energy⁽¹⁾ as opposed to significant losses reported by Husky Energy in 2020. The results also benefitted from favourable currency translation impacts, as well as the net impact of one-off items in 2021 as compared to those in 2020. These improvements were partly offset by lower contributions from the Telecommunications division as the operating environment remains challenging, particularly in Italy. In local currencies, the Group's reported EBITDA and EBIT grew 10% and 16% respectively from last year.

Impact on words vs visual

Analyses by Core Business Segments

	Pre-IFRS 16 ⁽¹⁾ 2021		Pre-IFRS 16 ⁽¹⁾⁽²⁾ 2020		Change	Local currencies change
	HK\$ million	%	HK\$ million	%	%	%
Revenue⁽³⁾						
Ports and Related Services ⁽³⁾	42,285	9%	32,865	8%	29%	26%
Retail	173,601	39%	159,619	40%	9%	6%
Infrastructure	56,100	13%	52,792	13%	6%	-
CK Hutchison Group Telecom	92,575	21%	90,663	22%	2%	-3%
Hutchison Asia Telecommunications	8,786	2%	9,147	2%	-4%	-5%
Finance & Investments and Others	72,036	16%	58,760	15%	23%	20%
Total Revenue	445,383	100%	403,846	100%	10%	7%
EBITDA⁽³⁾						
Ports and Related Services ⁽³⁾	15,157	13%	10,914	12%	39%	35%
Retail	16,034	14%	14,397	15%	11%	9%
Infrastructure	29,636	27%	29,066	30%	2%	-5%
CK Hutchison Group Telecom	43,052	39%	48,540	50%	-11%	-14%
Hutchison Asia Telecommunications	2,036	2%	2,034	2%	-	-1%
Finance & Investments and Others	5,312	5%	(8,007)	-9%	166%	165%
Total EBITDA	111,227	100%	96,944	100%	15%	10%

Impact on words vs visual



- Reported
- Underlying
- Europe
- Mainland China
- Hong Kong
- Asia, Australia & Others⁽¹⁾
- Canada
- Finance & Investments and Others

- Note 1: Includes Panama, Mexico and the Middle East
- Note 2: Prepared under Pre-IFRS 16 basis which is set out in note 1 on page 5
- Note 3: The outer pie chart represents EBITDA and EBIT %-mix on a reported basis. The inner pie chart represents underlying EBITDA and EBIT %-mix, which excludes the gain on disposal of tower assets completed in 2021 of HK\$25.3 billion, non-cash impairment of goodwill of the Group's Italian telecommunication business of HK\$(15.5) billion, impairment charge of Cenovus of HK\$(1.5) billion and a non-cash foreign exchange reserve loss of HK\$(3.5) billion following the Cenovus-Husky merger
- Note 4: Represents EBITDA and EBIT %-mix for Europe on a reported basis
- Note 5: Represents EBITDA and EBIT %-mix for Europe on an underlying basis

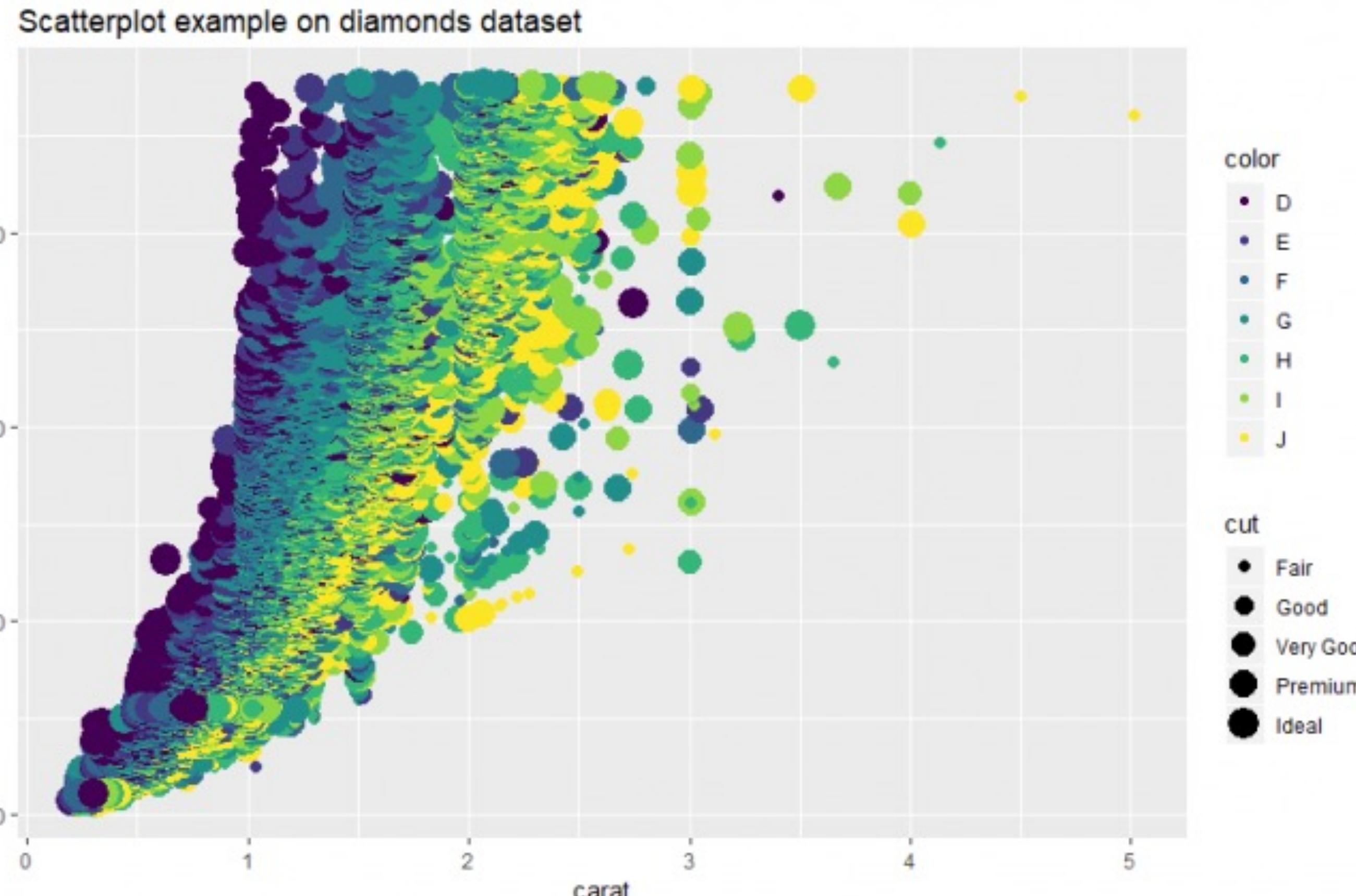
- Reported
- Underlying
- Ports & Related Services
- Retail
- Infrastructure
- Telecommunications
- Finance & Investments and Others

Visualization in Data Analysis

In data analysis and data science field, the volume of data is massive. We have to demonstrate the result in **humanized ways**.

Univ	Course Name	MS	DMIN	OMA	PROG	CAP	STOR	DMGT	SEC	BIG	PREP	GOV	DEC	COMM	VIS	ETH	CASE	EVAL
CC	Data Mining	x	x	x	x		x			x								
	Introduction to Data Science	x	x	x						x								
	Dataset Organization and Management					x	x	x		x								
	Data Science Capstone	x						x									x	
NKU	Data Visualization														x			
	Data Mining	x	x		x													x
	Big Data					x	x		x									
	Intro to Data Science								x									
	Data Analytics			x								x	x					
OSU	Data Science Capstone			x						x			x	x				
	Intro to Scientific Visualization				x					x			x					
	Introduction to Data Mining	x	x	x						x			x					
UR	Introduction to Data Mining	x	x							x			x			x		
USF	Data Visualization			x	x				x			x				x		
	Data Mining	x	x	x	x				x	x								
	Mathematical Modeling		x															x

Table 10: Mapping of competencies to courses in data science programs



Visualization key principal

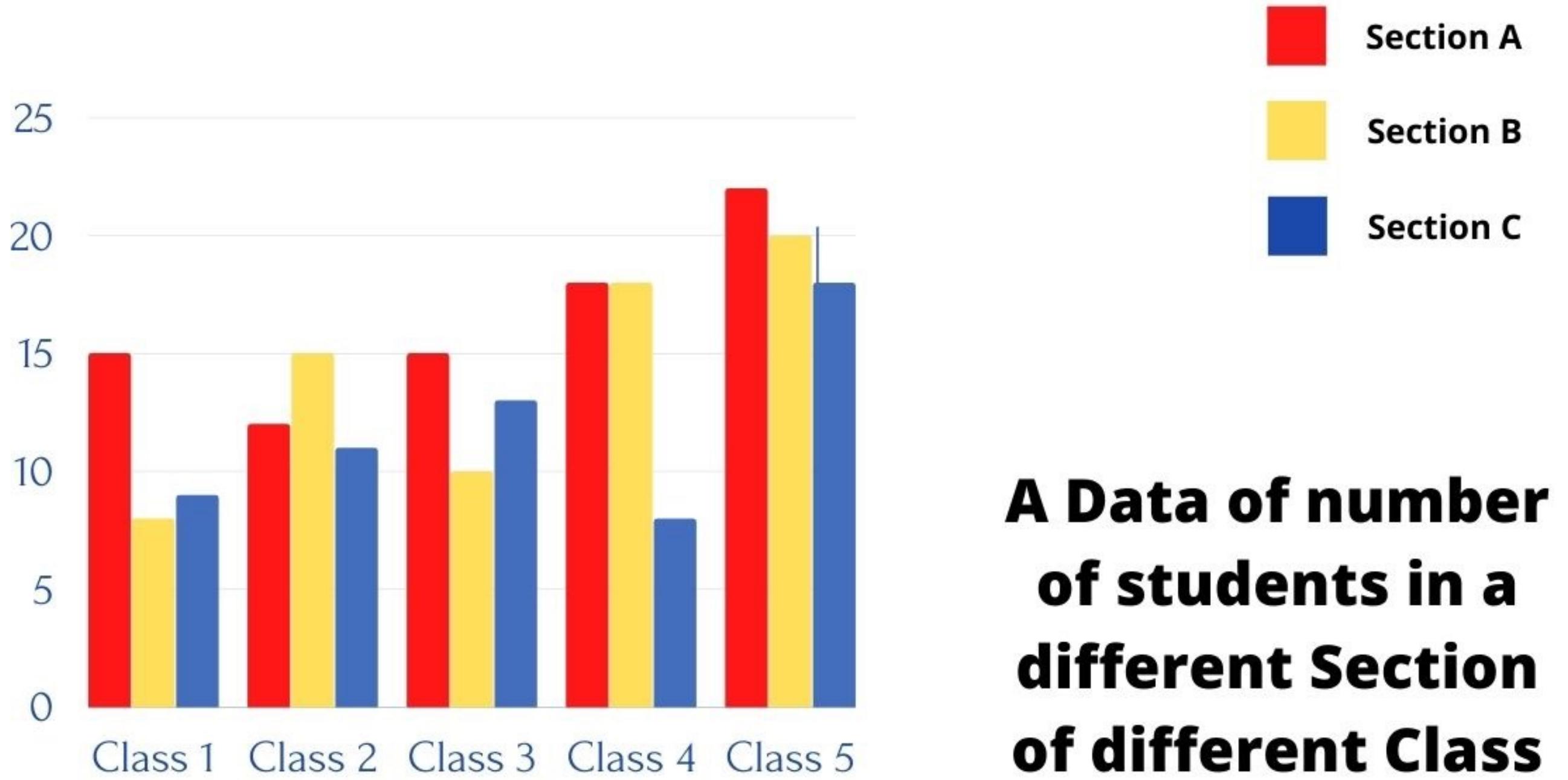
Key principles of effective data visualization

- Determine the best visual
- Balance the design
- Focus on the key areas
- Keep it simple
- Use patterns
- Compare aspects



Common types of visual chart

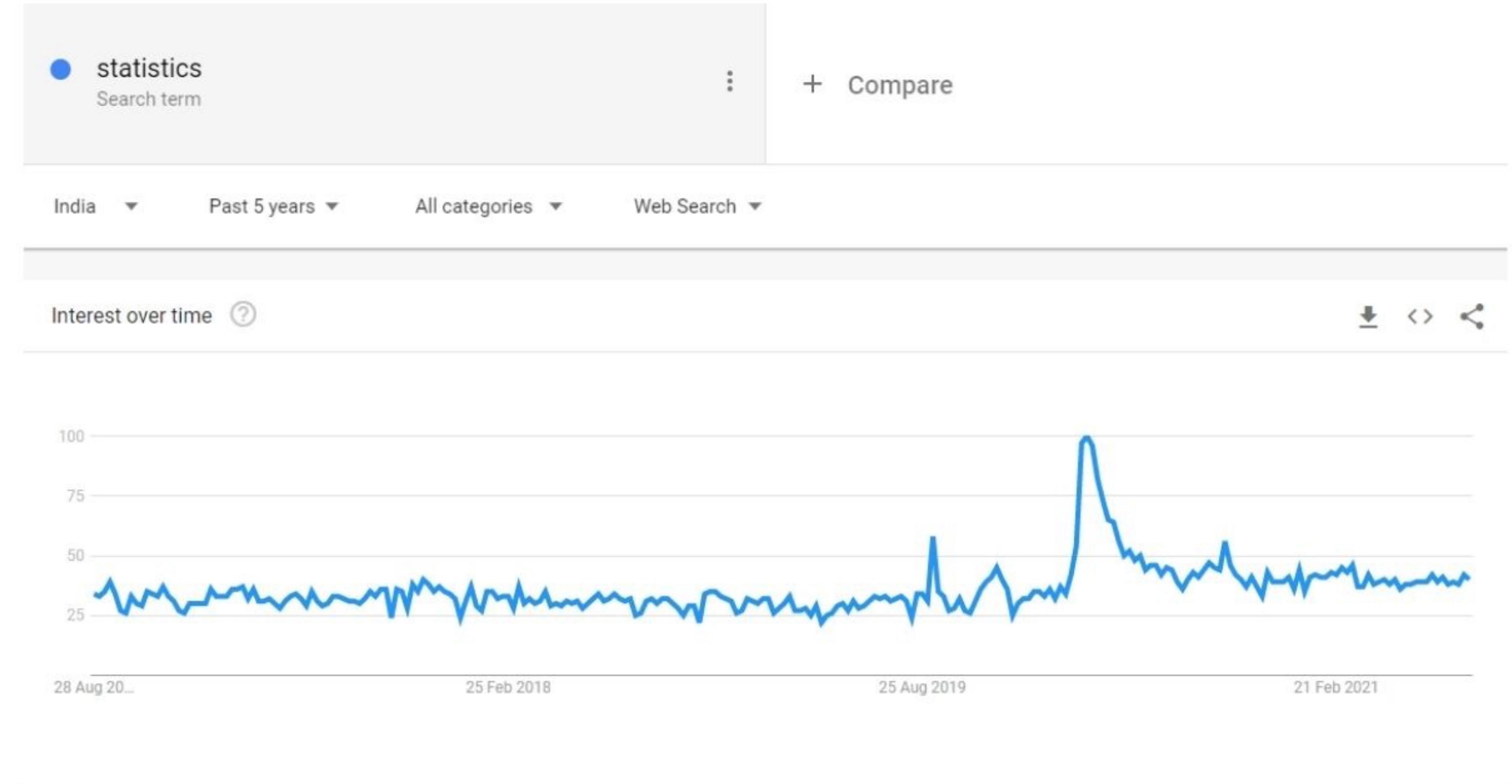
Bar chart used to swiftly compare data across categories, emphasize discrepancies, disclose historical highs and lows, and indicate trends and outliers.



**A Data of number
of students in a
different Section
of different Class**

Common types of visual chart

Line chart, also known as a line graph, are used to visualize data trends, usually across time (like stock price changes over five years or website page views for the month)



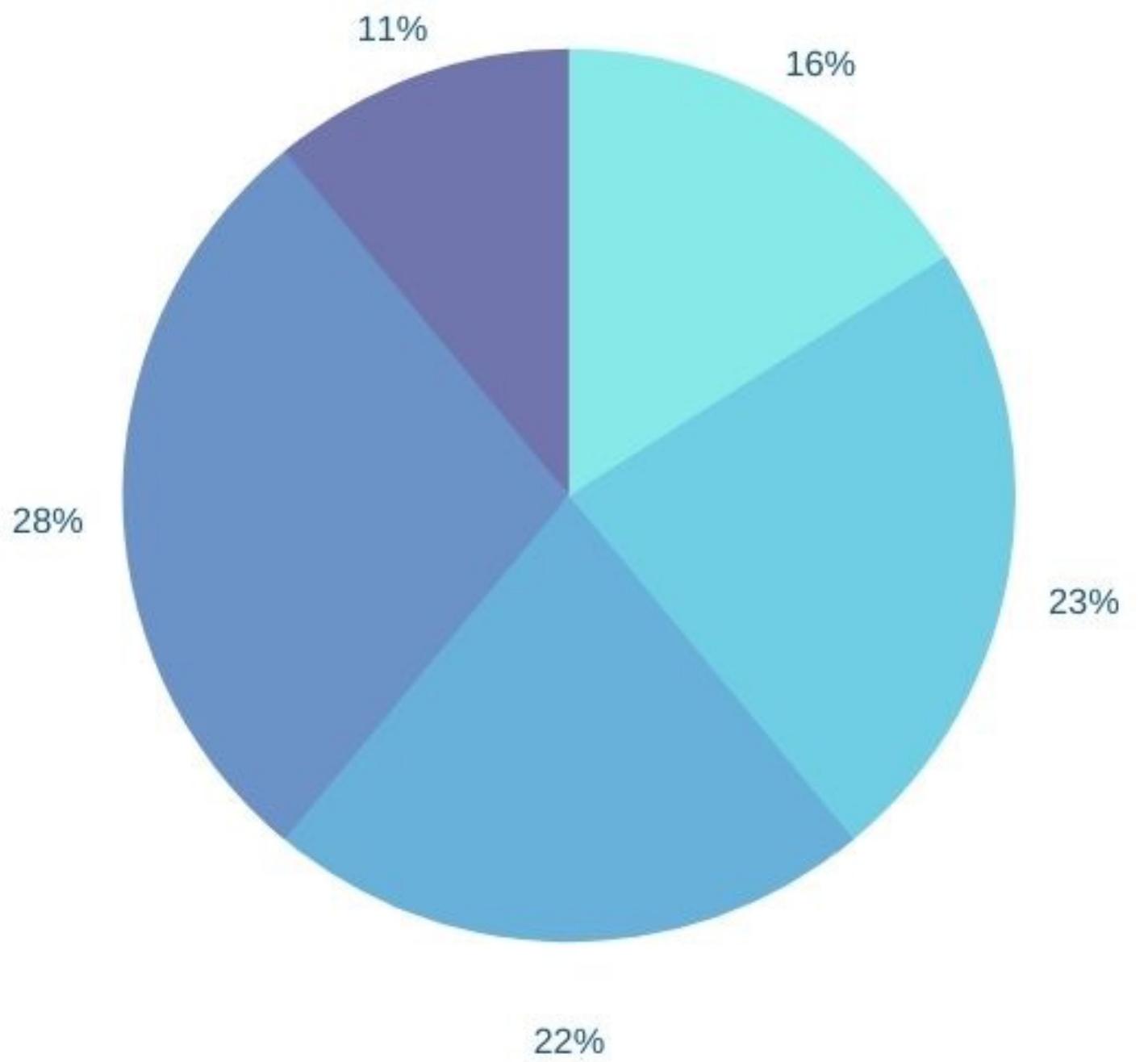
Common types of visual chart

Pie chart is a circular representation of the statistical graphic, which is divided into various slices to show all the desired data in numerical proportions.

Website Visits

Site Traffic Sources

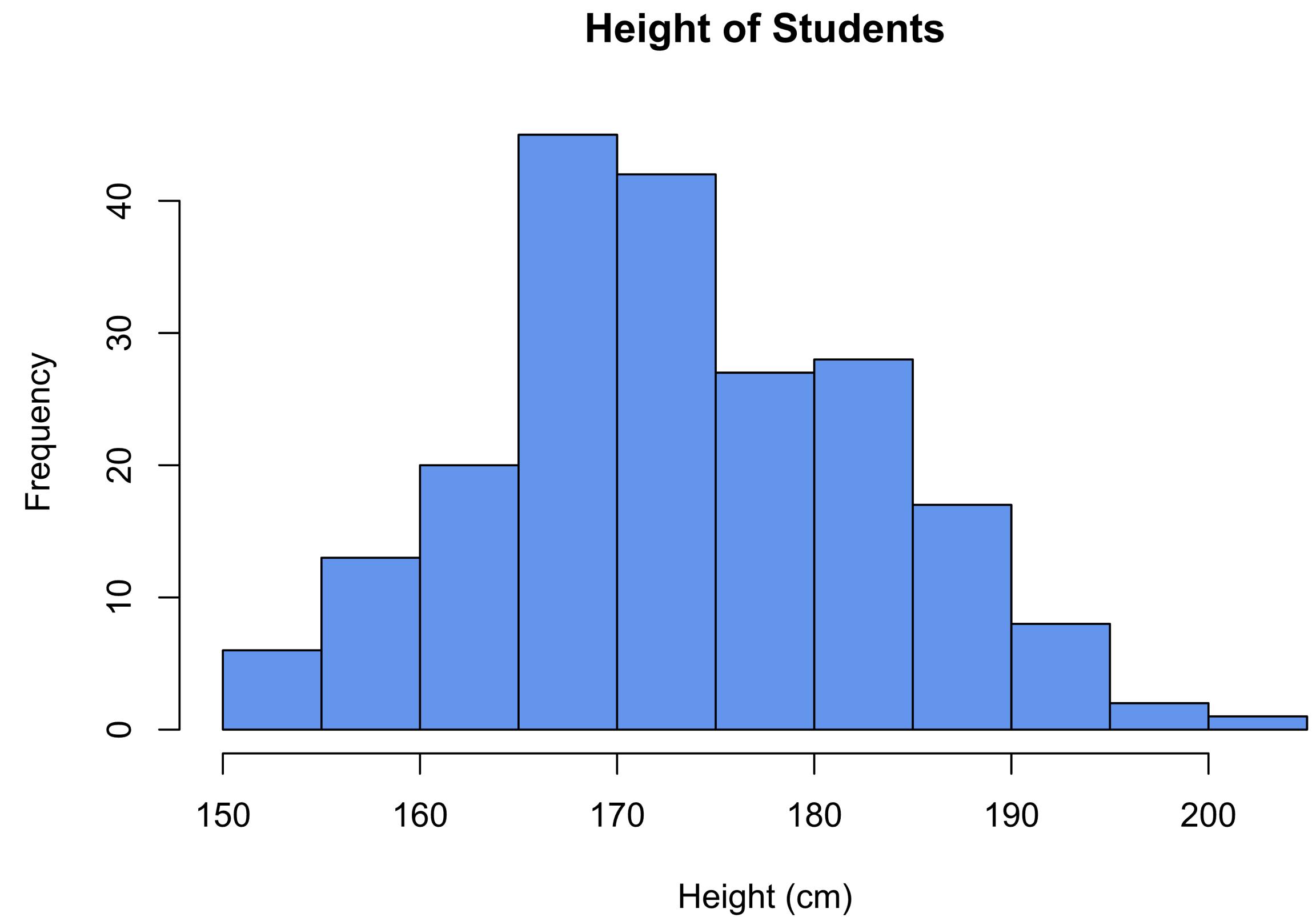
- ORGANIC
- SEARCH ENGINE
- EMAIL MARKETING
- SOCIAL MEDIA
- REFERRAL LINK



Common types of visual chart

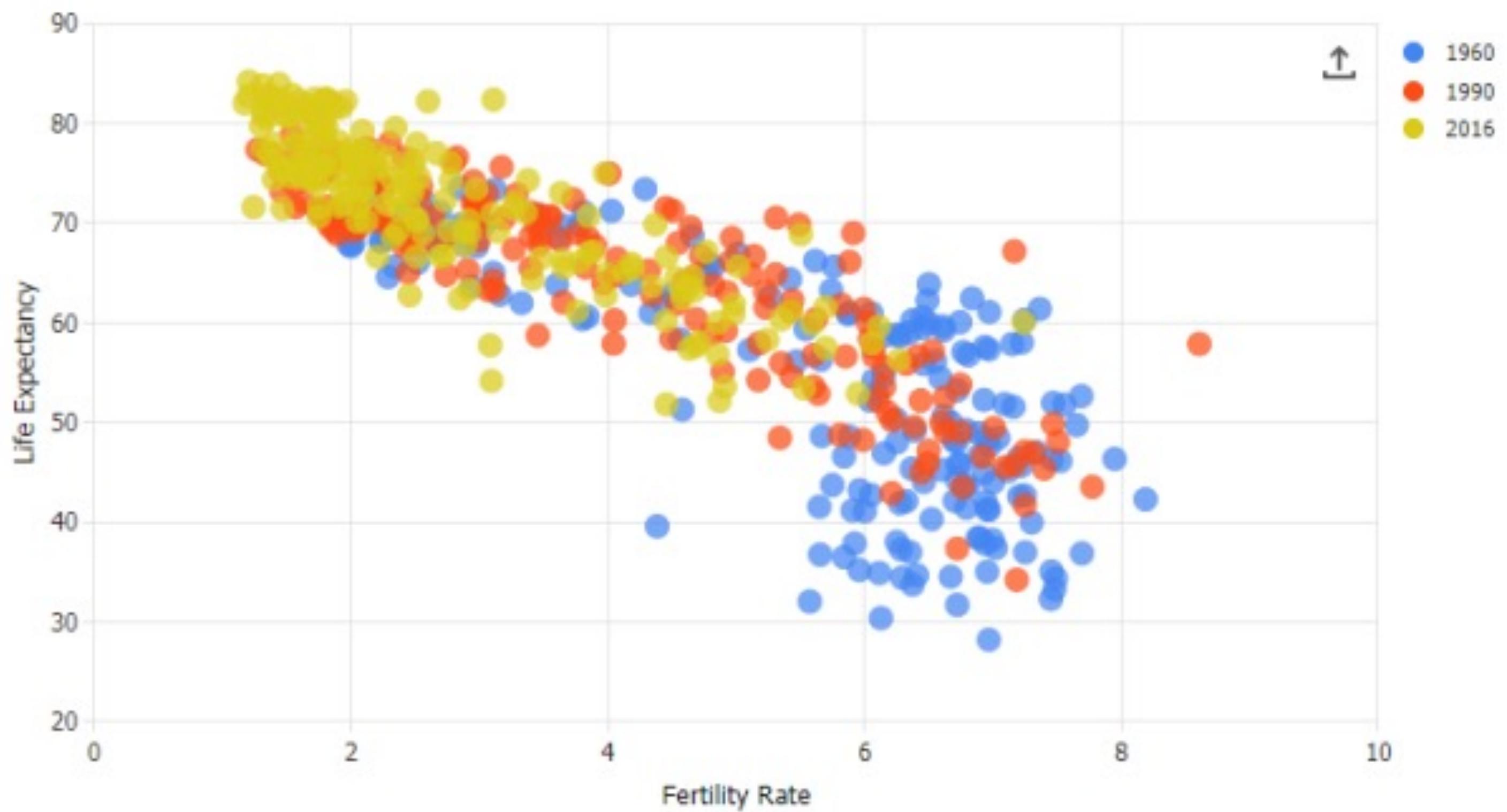
Histogram used to summarise data that are either discrete or continuous and measured on an interval scale.

It is frequently used to visualize the significant features of the data distribution in a convenient format.



Common types of visual chart

Scatter plots are helpful in examining the relationship between two or more variables, revealing whether one is a good predictor of the other or whether they tend to vary independently.



Visualization tools in data analysis



Matplotlib

Matplotlib is the result of development efforts by John Hunter (1968–2012) and the project's many contributors.

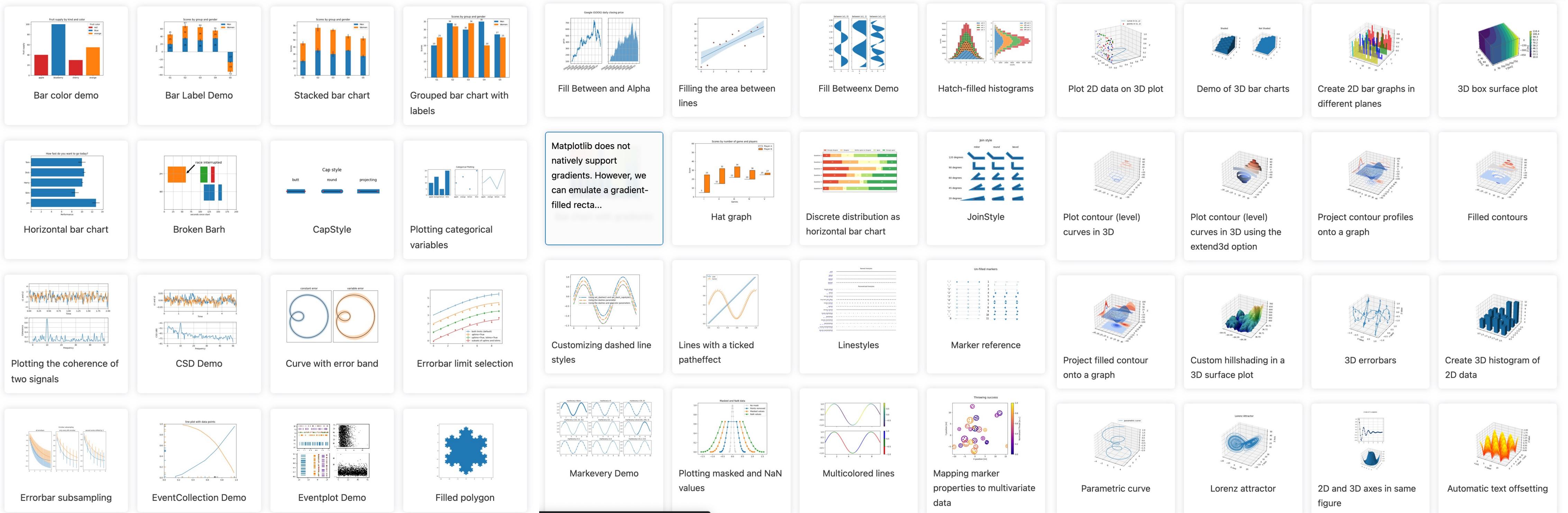
Matplotlib is a community project maintained for and by its users.

Matplotlib is the most common tools in data analysis and data science industries.



Matplotlib

Matplotlib is a comprehensive visualization tools work with iPython.



-gallery-lines-bars-and-markers-gradient-bar-py

Matplotlib

Bar chart example by Matplotlib

```
import matplotlib.pyplot as plt

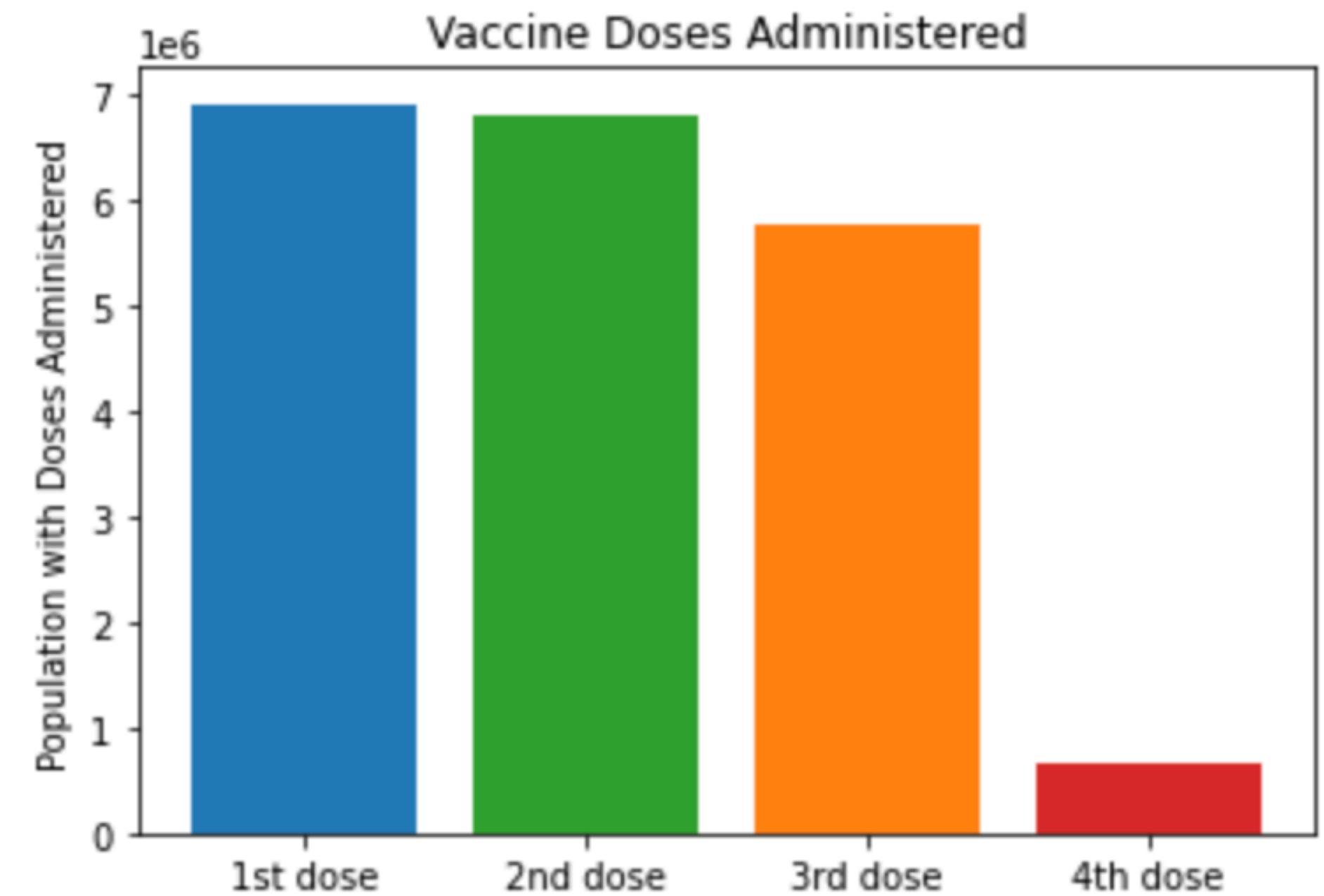
fig, ax = plt.subplots()

vaccine = ['1st dose', '2nd dose', '3rd dose', '4th dose']
counts = [6904243, 6779247, 5764953, 675118]
bar_colors = ['tab:blue', 'tab:green', 'tab:orange', 'tab:red']

ax.bar(vaccine, counts, label=bar_labels, color=bar_colors)

ax.set_ylabel('Population with Doses Administered')
ax.set_title('Vaccine Doses Administered')

plt.show()
```



Seaborn



seaborn

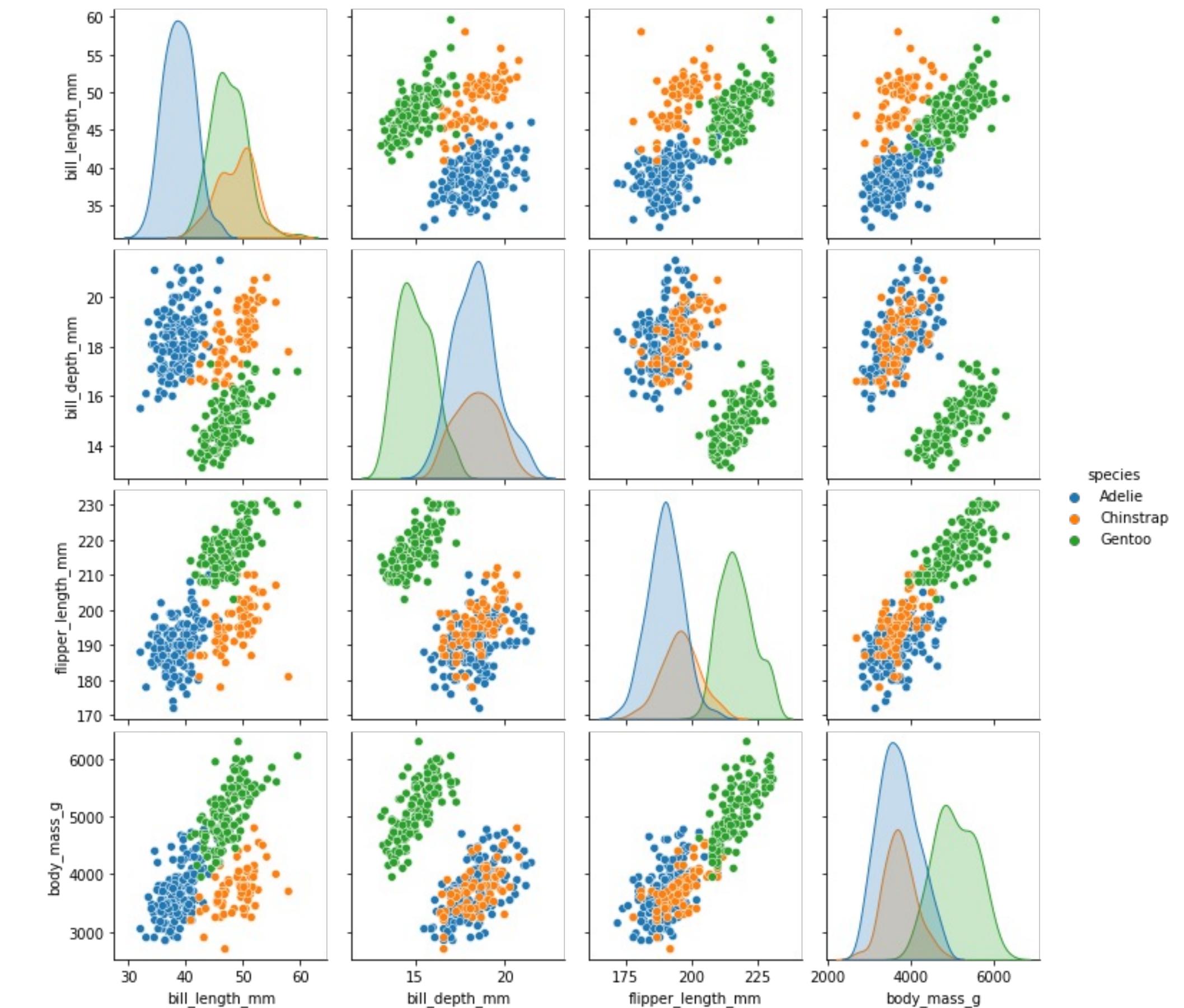
Seaborn is a high level interface for drawing statistical graphics **with Matplotlib**. It aims to make visualization a central part of exploring and understanding complex datasets.

You may treat Seaborn as a plug-in for Matplotlib to make ease of coding and design.

Seaborn

Seaborn plot better designed graph in simple code base on Matplotlib origin, meaning you must have installed both before use.

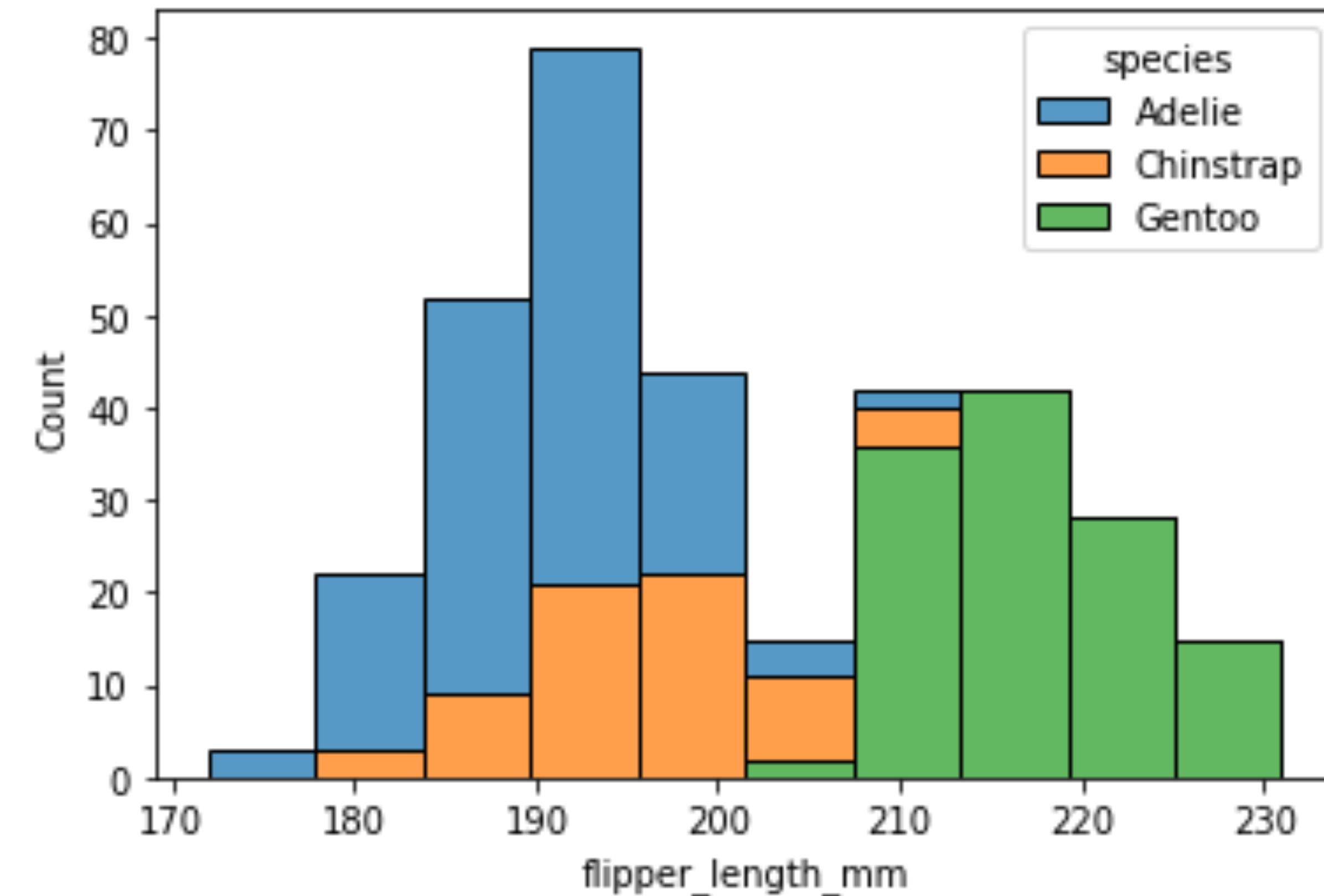
```
1 | sns.pairplot(penguins, hue="species")
```



Seaborn – simple coding style

```
1 sns.histplot(data=penguins, x="flipper_length_mm", hue="species", multiple="stack")
```

```
<AxesSubplot:xlabel='flipper_length_mm', ylabel='Count'>
```

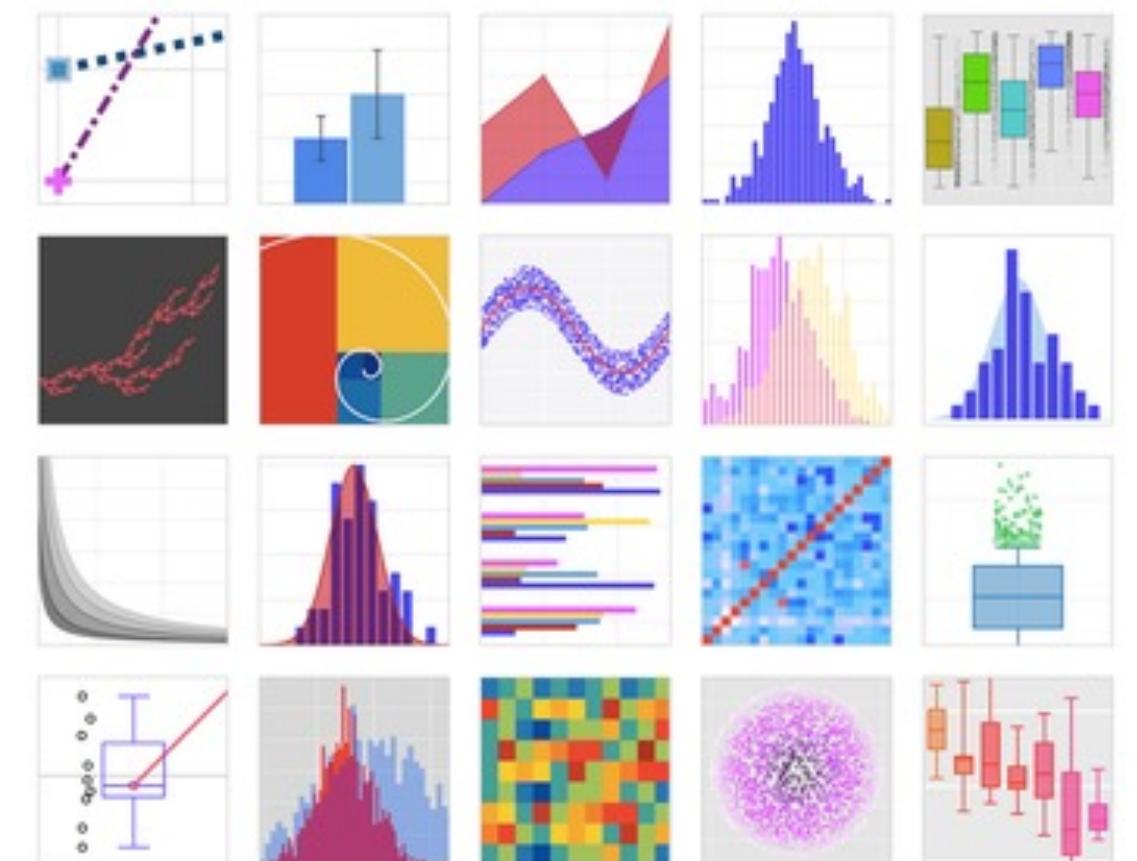


Plotly

Plotly was founded in 2012, relatively fresher than Matplotlib. It develops online/offline data analytics and visualization tools.

Plotly provides online graphing, analytics, and statistics tools for individuals and collaboration, as well as scientific graphing libraries for Python, R, MATLAB, Perl, Julia, Arduino, and REST.

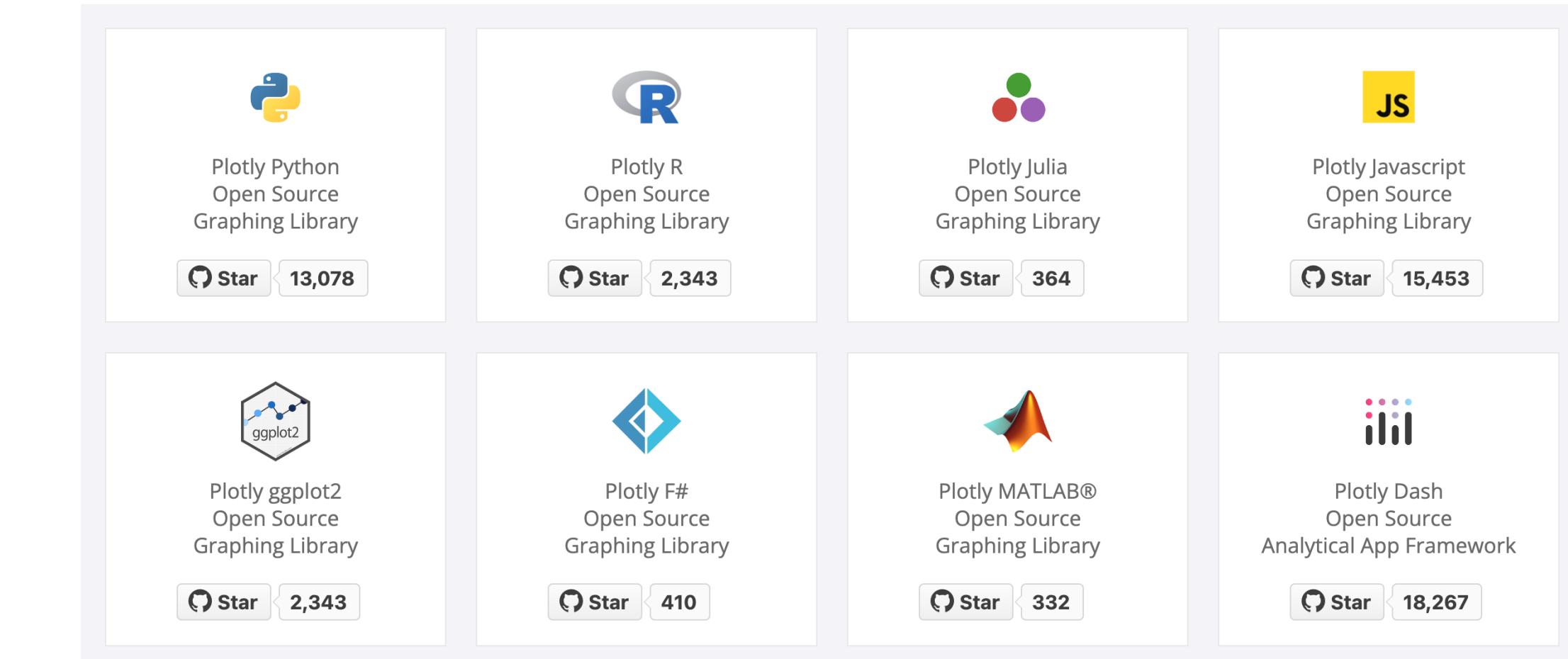
As its simple coding feature and interactive well designed graphs visualization, the user population is growing fast!



Plotly

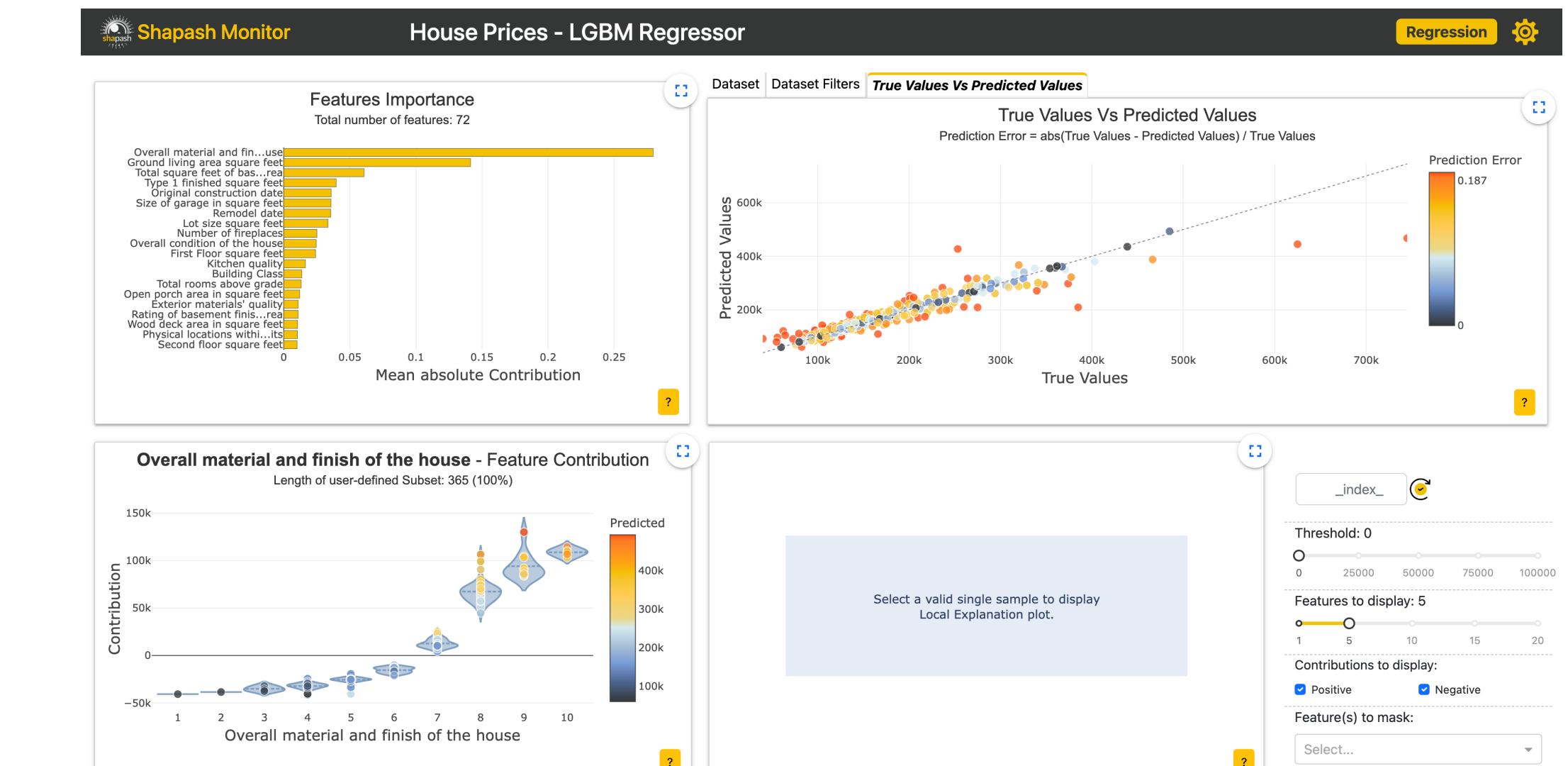


Plotly's Python graphing library **makes interactive, publication-quality** graphs. In this chapter we are focusing this module for data analysis.



Dash is the original **low-code framework** for rapidly **building data apps** in Python, R, Julia, and F# (experimental).

Written on top of Plotly.js and React.js, Dash is ideal for building and deploying data apps with customized user interfaces. It's particularly suited for anyone who works with data.



Plotly – graphing library

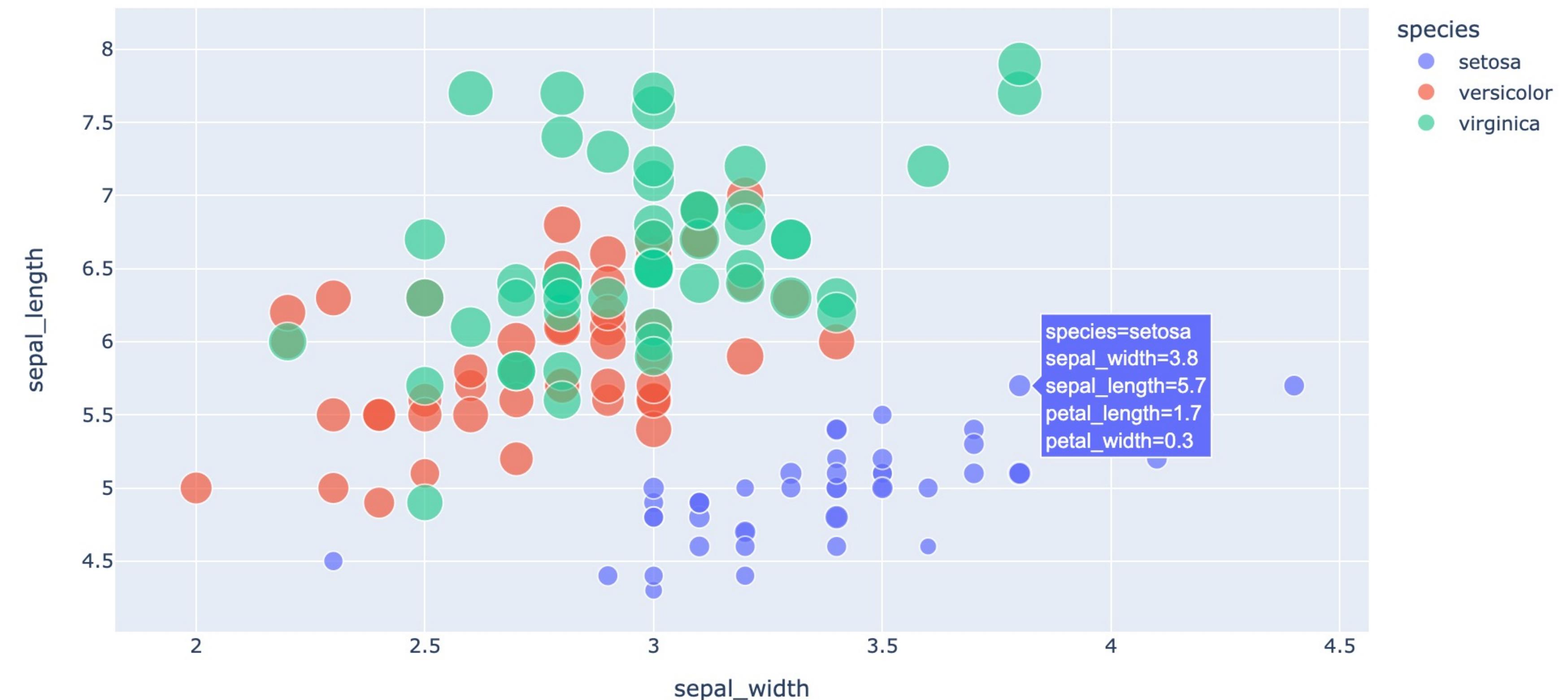
```

1 fig = px.scatter(df, x="sepal_width", y="sepal_length", color="species",
2                   size='petal_length', hover_data=['petal_width'])
3 fig.show()

```



One of the most handy benefit of plotly is its **interactive** data on **hovering**.

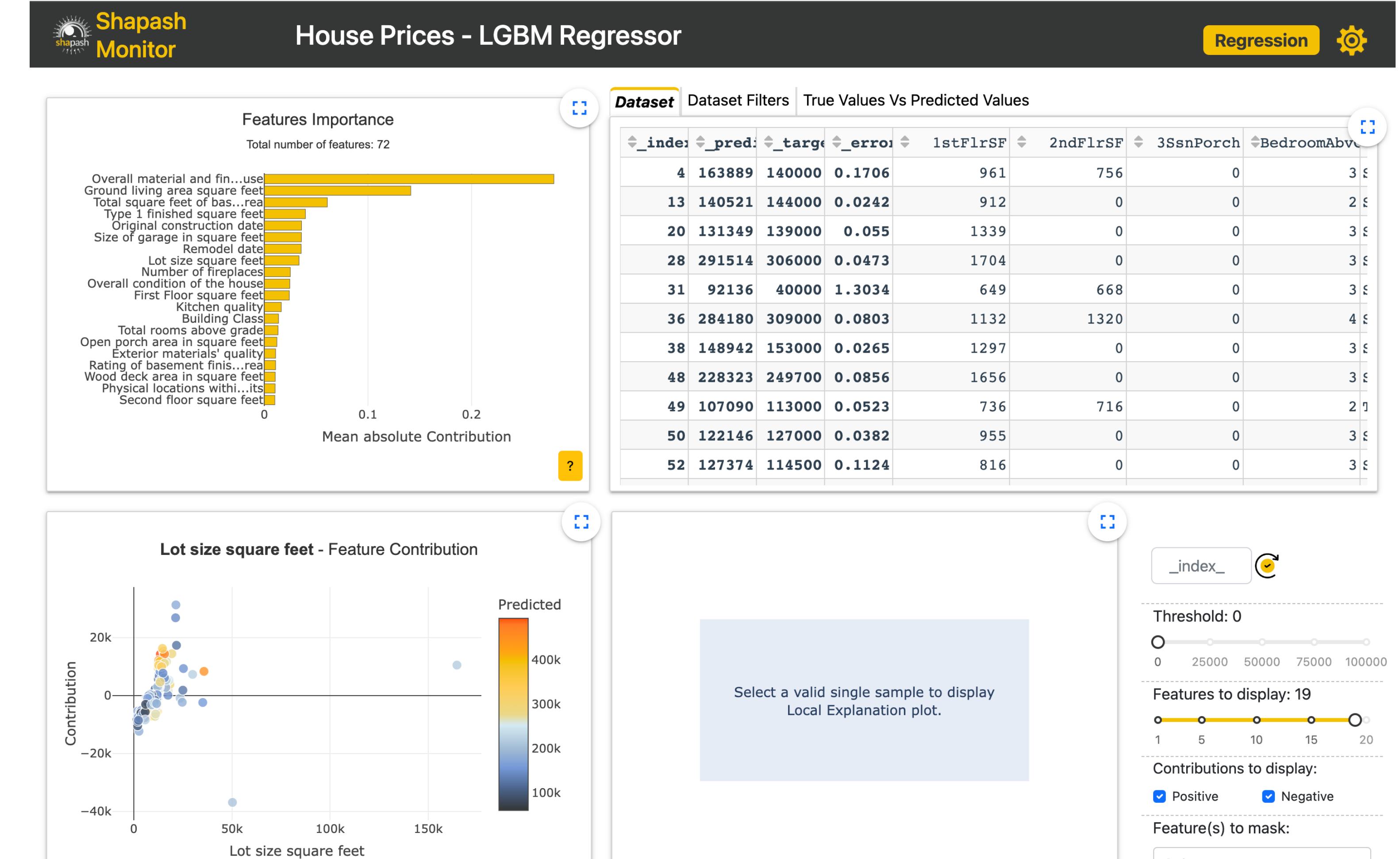


Plotly Dash

 **plotly | Dash**
 **plotly | Dash Enterprise**

is built for website
 interactive data
 visualization.

Demo link



Tableau

Tableau is a leading data visualization tool used for data analysis and business intelligence.

It is also a dashboard environment visual tool.



Tableau Features

- Tableau supports powerful data discovery and exploration that enables users to answer important questions in seconds
- No prior programming knowledge is needed; users without relevant experience can start immediately with creating visualizations using Tableau
- It can connect to several data sources that other BI tools do not support. Tableau enables users to create reports by joining and blending different datasets
- Tableau Server supports a centralized location to manage all published data sources within an organization

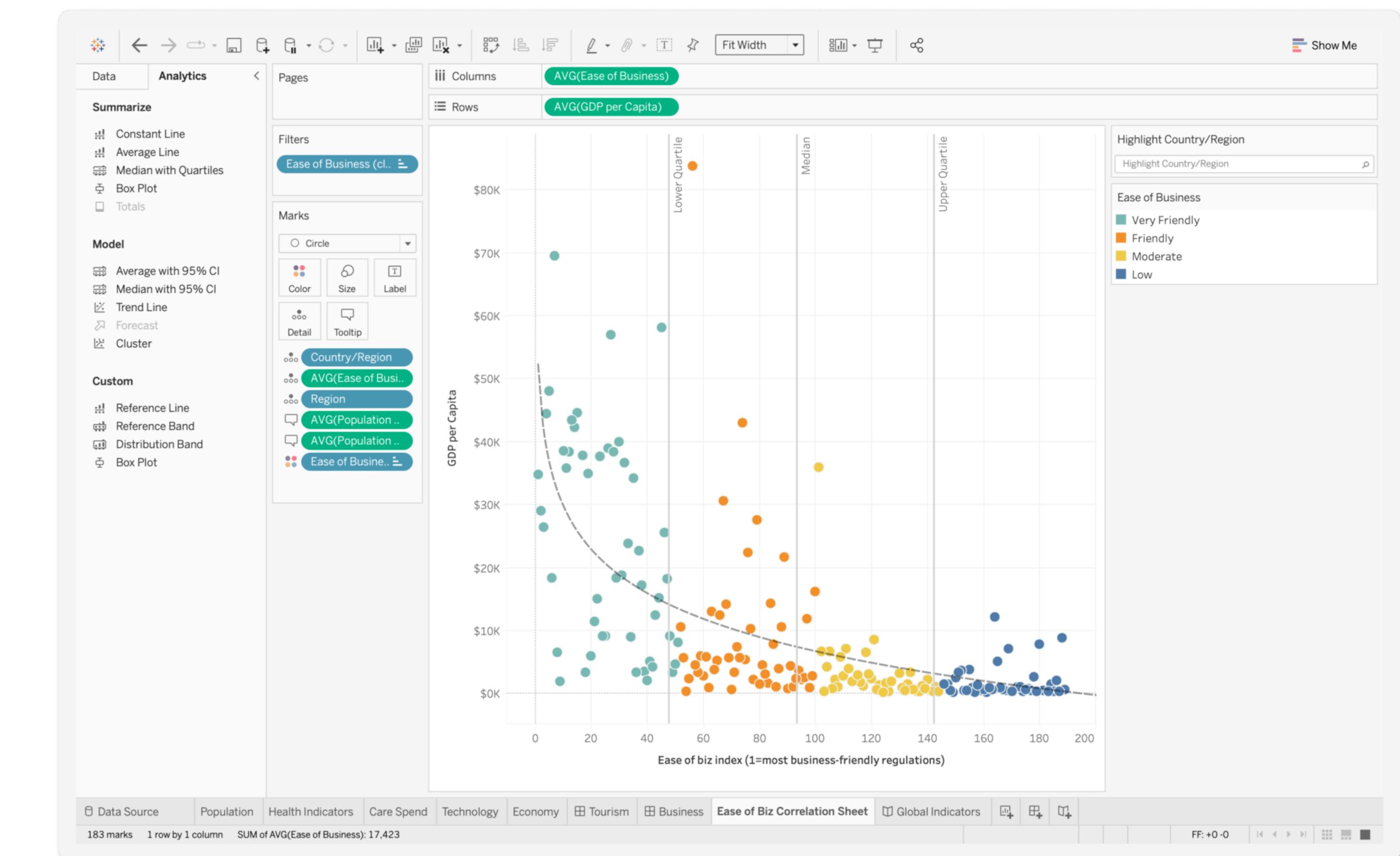
Tableau

Tableau drives better business outcomes with fully-integrated data management and governance, visual analytics and data storytelling, and collaboration—all with Salesforce's industry-leading built right in.



Tableau

Tableau Desktop delivers to access, visualize, and analyse your data. With an intuitive drag and drop interface, you can uncover the hidden insights to make impactful business decisions faster.



Apache Superset

Superset is fast, lightweight, intuitive, and loaded with options that make it easy for users of all skill sets to explore and visualize their data, from simple line charts to highly detailed geospatial charts.

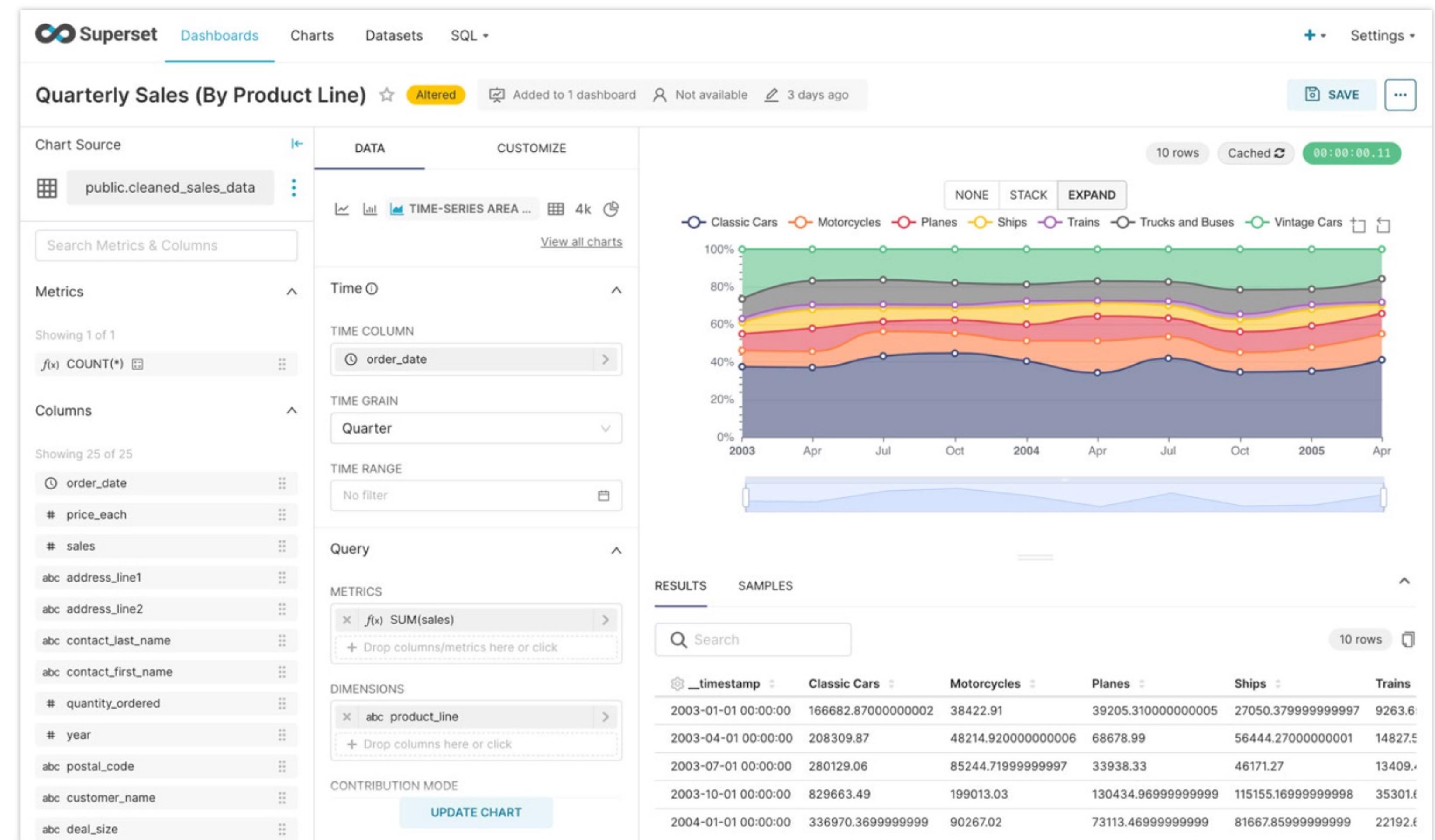


Features:

- Dashboard creation
- Enterprise authentication ([OpenID](#), [LDAP](#), [OAuth...](#))
- Integration with [Apache ECharts](#)
- Lightweight [semantic layer](#)
- Visualization plugin support
- Compatible with most [SQL](#)-speaking data sources

Superset

- Explore your data using the array of data visualizations.
- View your data through interactive dashboards
- Use SQL Lab to write queries to explore your data



Superset

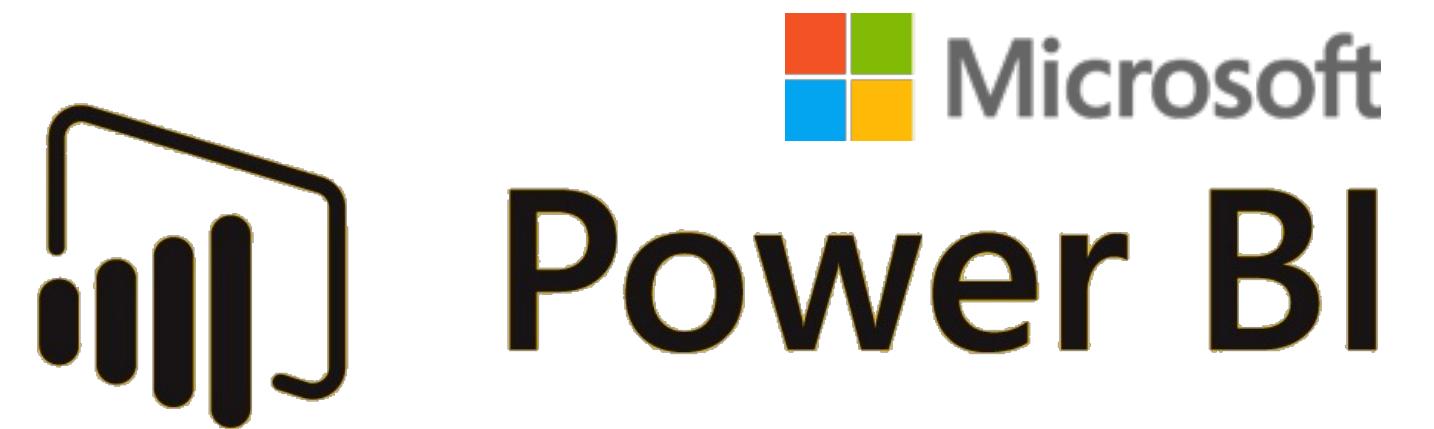
Supported Databases



... and many other compatible databases

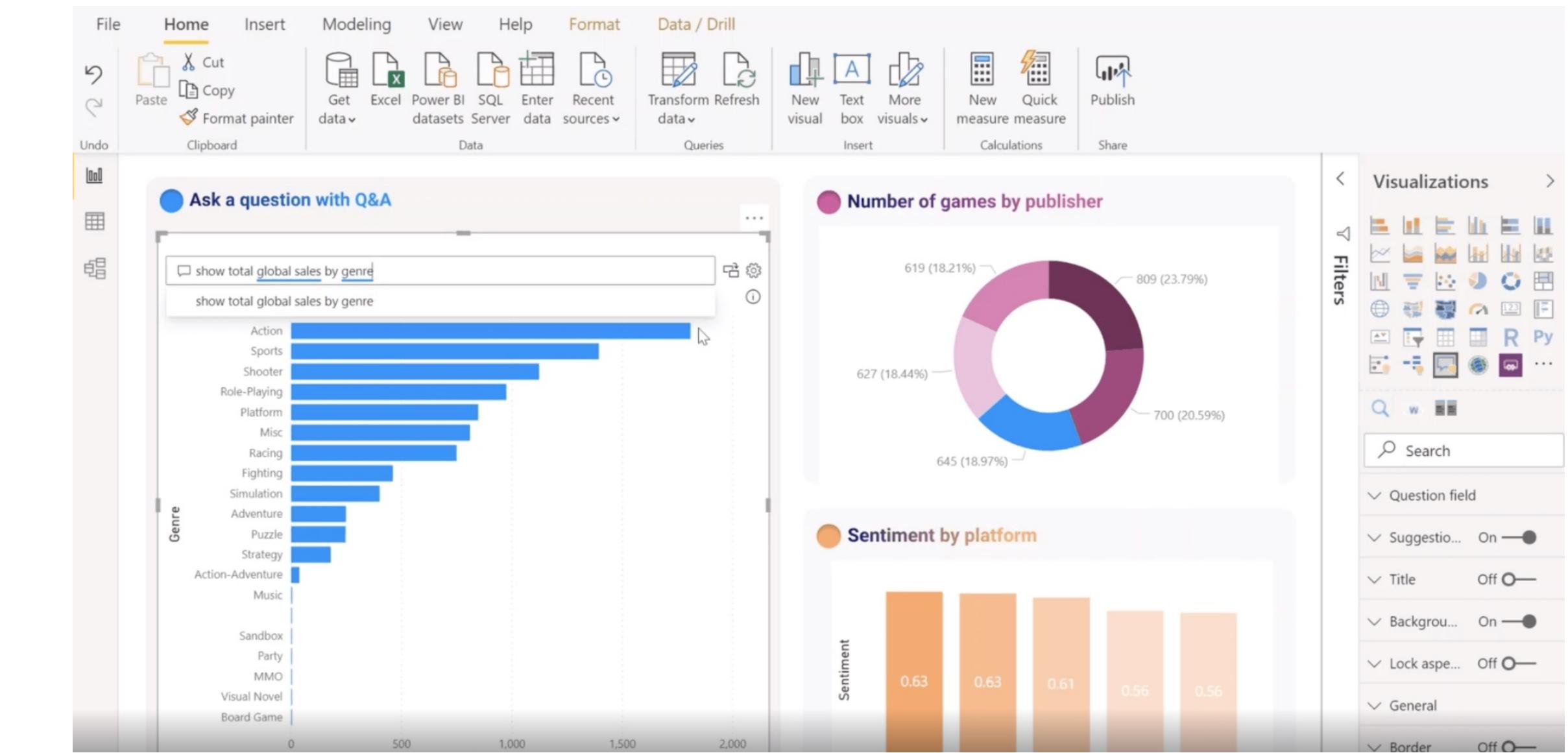
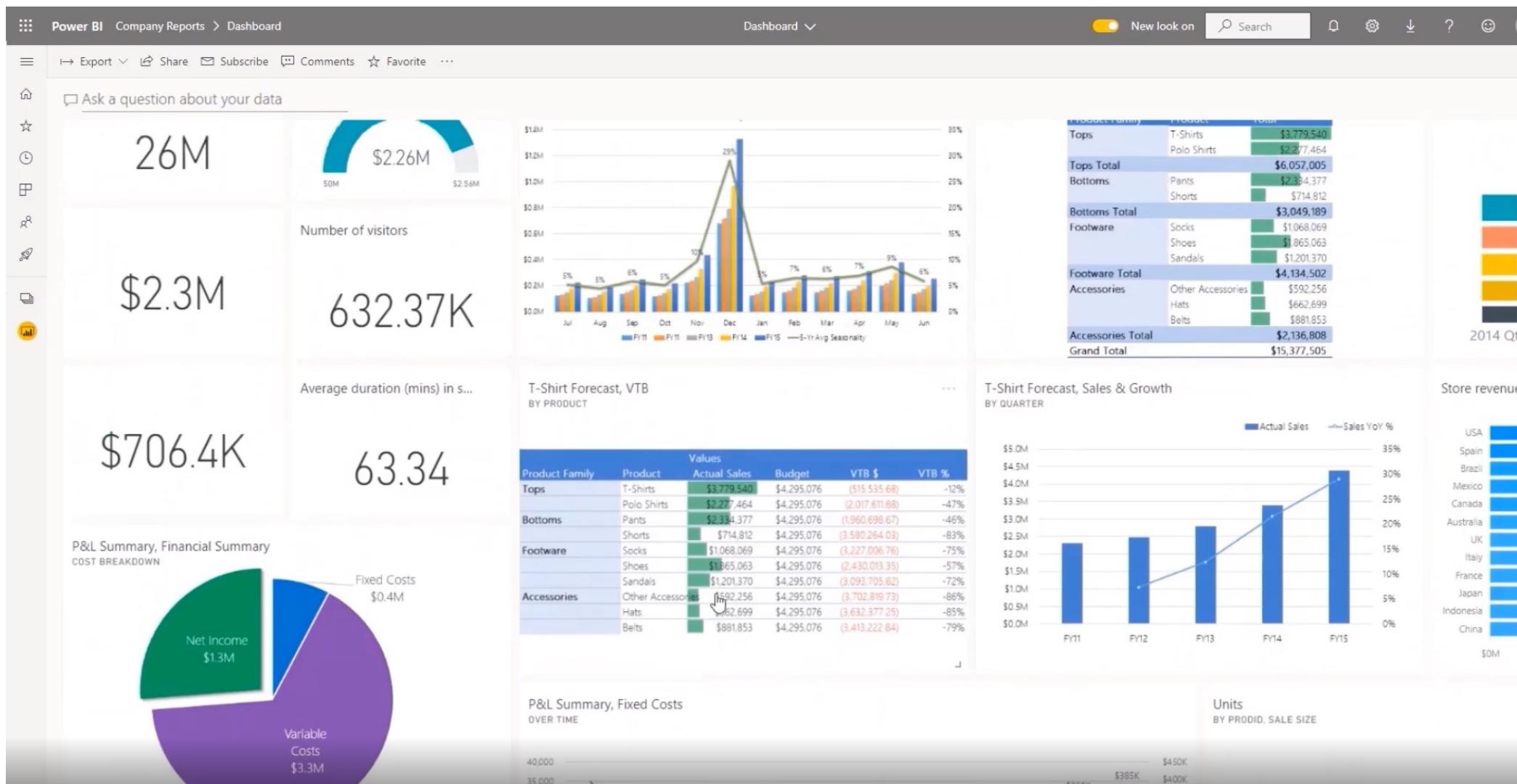
Power BI

Power BI created by Microsoft. Power BI is a unified, scalable platform for self-service and enterprise **business intelligence** (BI). Connect to and visualise any data, and infuse the visuals into the apps.



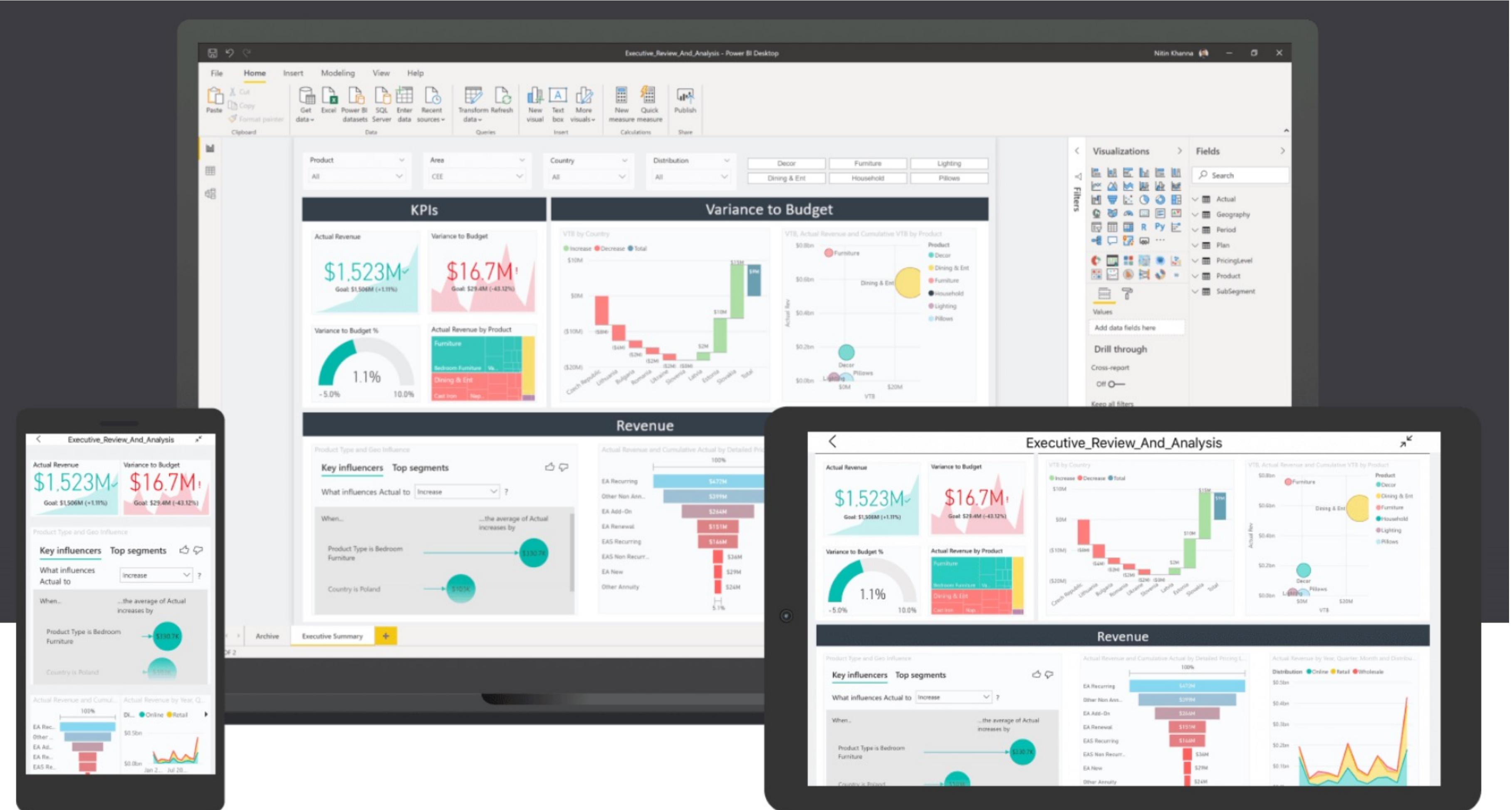
Power BI

Easily connect to, model, and visualise data, creating memorable reports personalised with your KPIs and brand. Get fast, AI-powered answers to business questions—even.



Work together easily on the same data, collaborate on reports, and share insights across popular Microsoft Office applications such as Microsoft Teams and Excel to quickly make data-driven decisions that drive strategic actions.

Power BI: msOffice-like interface



Summary

Work with Python



Stand alone



Reference



<https://matplotlib.org/>



<https://seaborn.pydata.org/>



<https://plotly.com/graphing-libraries/>



<https://dash.plotly.com/>



<https://www.tableau.com/>



<https://superset.apache.org/>



Power BI

<https://powerbi.microsoft.com/>

