

DATA AND ARTIFICIAL INTELLIGENCE



Introduction to Data Analytics



Data Visualization for Decision-Making

Learning Objectives

By the end of this lesson, you will be able to:

- 🕒 Explain data visualization
- 🕒 Describe the importance of data visualization
- 🕒 List various tools of data visualization



Data Visualization

Data Visualization



- Data visualization is the graphical representation of data using charts, graphs, and maps.
- Our eyes are drawn to colors and patterns.
- Data visualization is a form of visual art that grabs our interest and keeps our eyes on the message.

Data Visualization

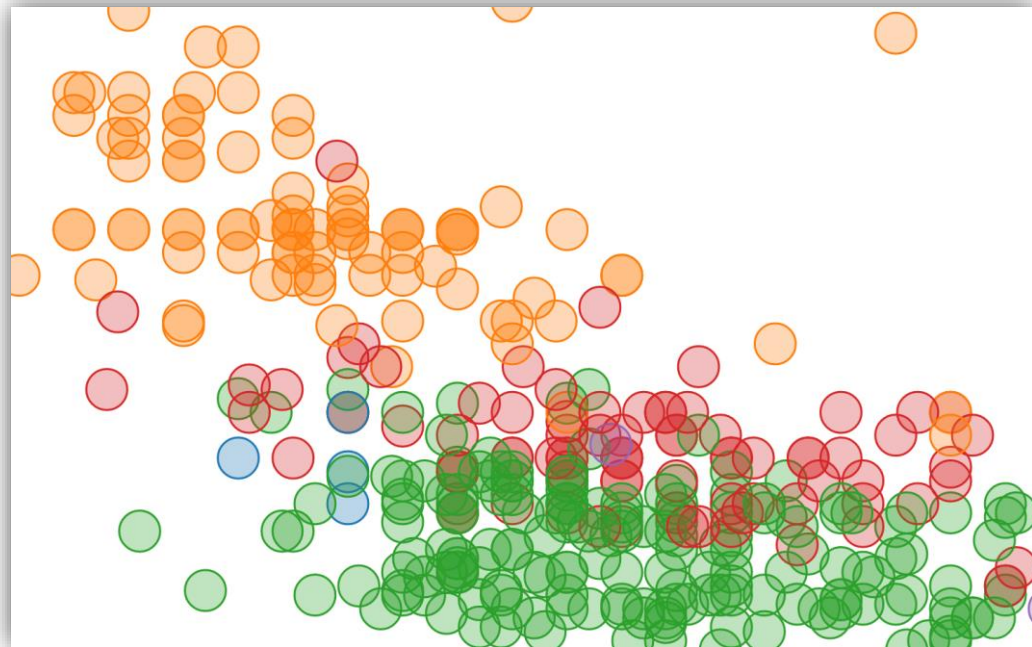


Visualized data is more effective and consumable than a massive spreadsheet of data.

Data Visualization

Year	JAN	FEB	MAR	APR	MAY
1910	111.4	126.1	49.9	95.3	71.8
1911	59.2	99.7	62.1	69.0	52.2
1912	111.7	79.5	128.2	36.1	58.2
1913	123.4	57.1	131.2	102.9	81.5
1914	78.8	114.9	124.3	52.3	59.6
1915	118.7	141.1	55.0	65.9	53.2
1916	108.9	140.3	88.3	70.2	93.0
1917	73.6	33.8	74.2	63.6	69.2
1918	105.2	106.3	42.9	49.6	65.0
1919	120.1	59.8	118.4	80.9	29.8
1920	139.7	96.1	109.6	106.1	94.4
1921	149.3	24.1	103.9	38.9	62.8
1922	124.1	114.7	72.4	86.1	63.4
1923	104.5	152.6	51.1	79.1	78.4

- The table shows the total sale of products corresponding with each year.

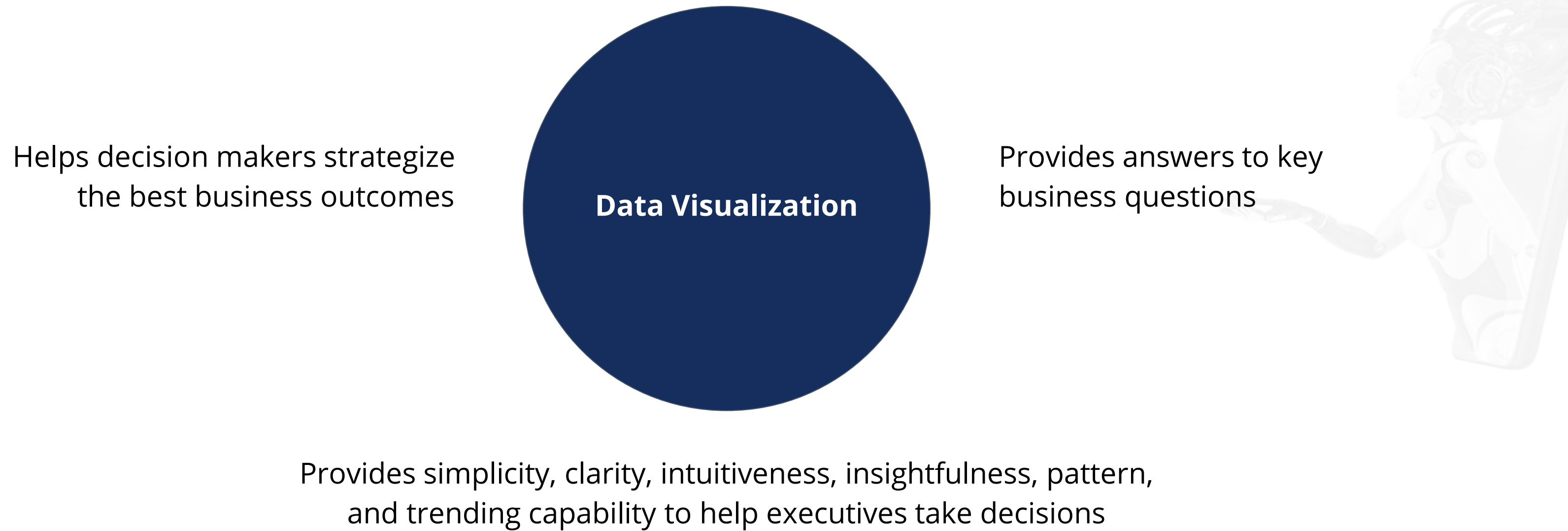


- The adjacent graph is the visualization of the sale data points using a visualization tool.

Understanding Data Visualization

Understanding Data Visualization

Data analytics allows decision makers and executives to weigh the alternatives of different outcomes of their decisions.



Benefits of Data Visualization

- Sales reports are formal documents or PowerPoint slides with many tables and charts.
- They are elaborate and the real point is lost in the data.
- Data visualization helps by making information crisp, clear, and memorable.



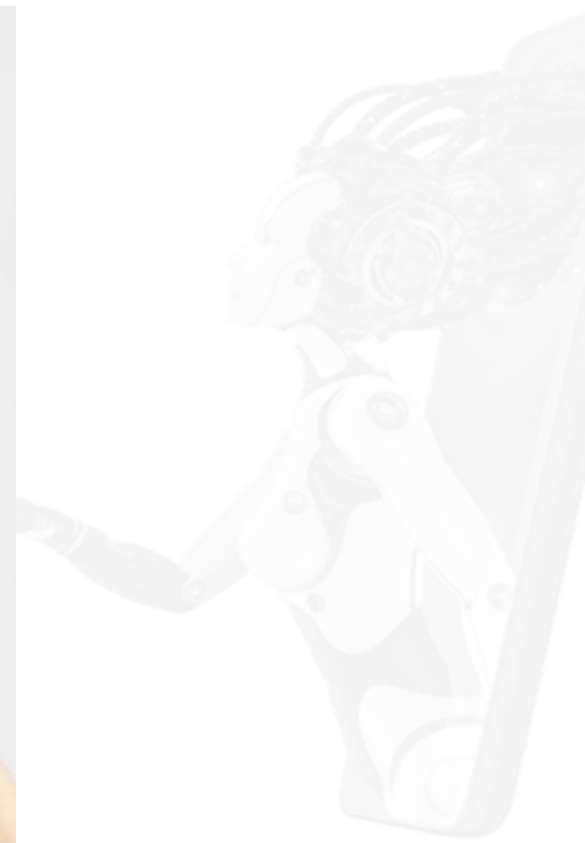
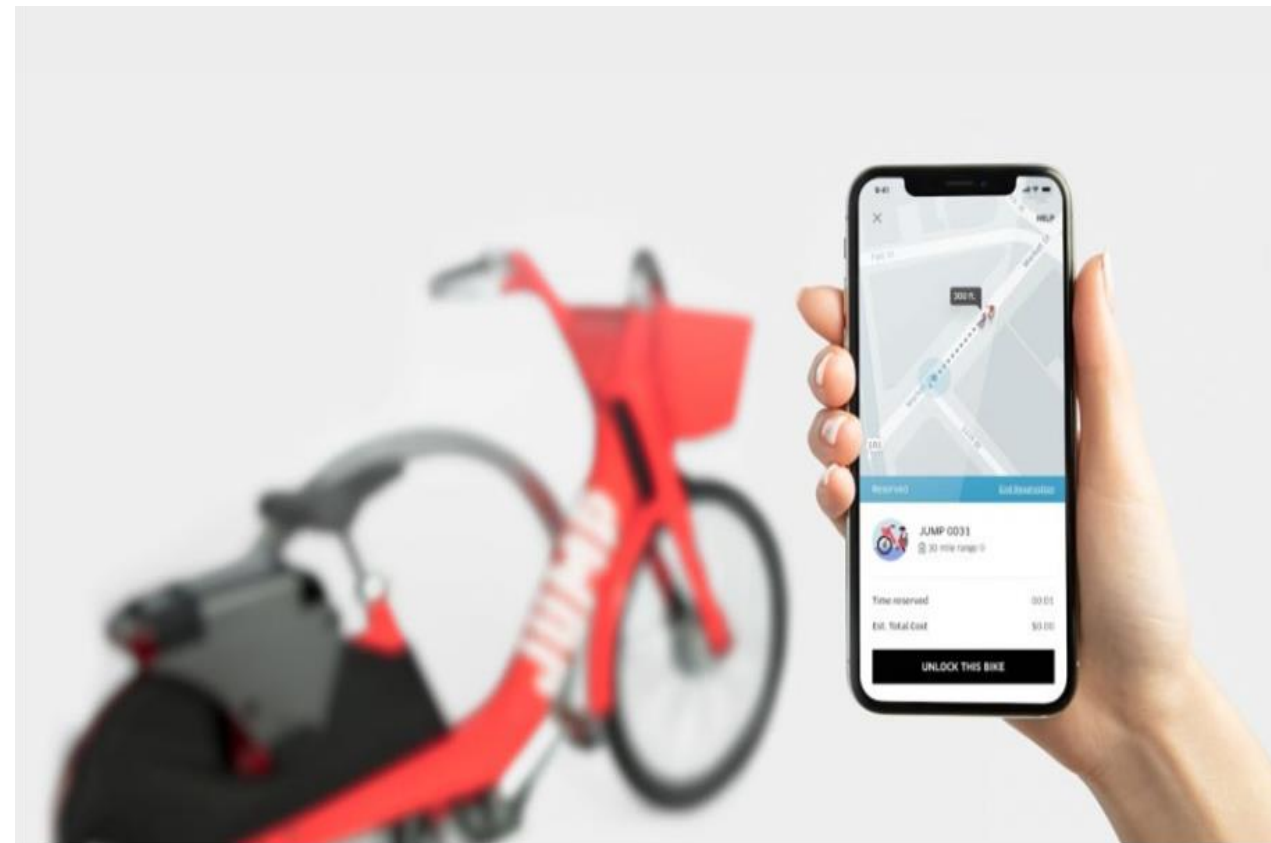
Benefits of Data Visualization: Example

- From a bar chart, a sales director can identify that the sales of their flagship product in the southwest region is going down by eight percent.
- The director can spot the occurrence of variances and start formulating a plan to improve the sale.
- Data visualization allows executives to spot problems and act on them.



Data Visualization: Uber

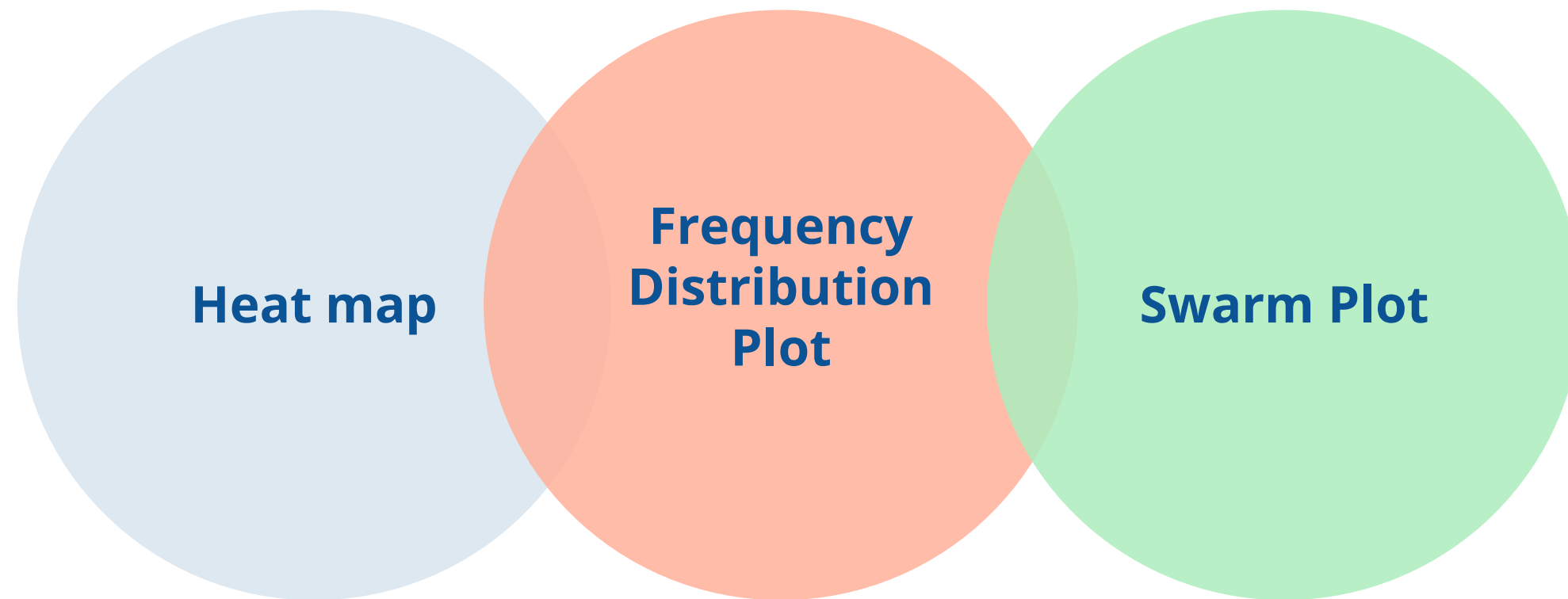
In October 2018, Uber released a visualization product that provides insights on mobility for JUMP bikes.



This data is helpful in urban areas to evaluate the success of their shared bike program. It can also help in planning infrastructure investments in cities to promote safety and smooth mobility.

Commonly Used Visualizations

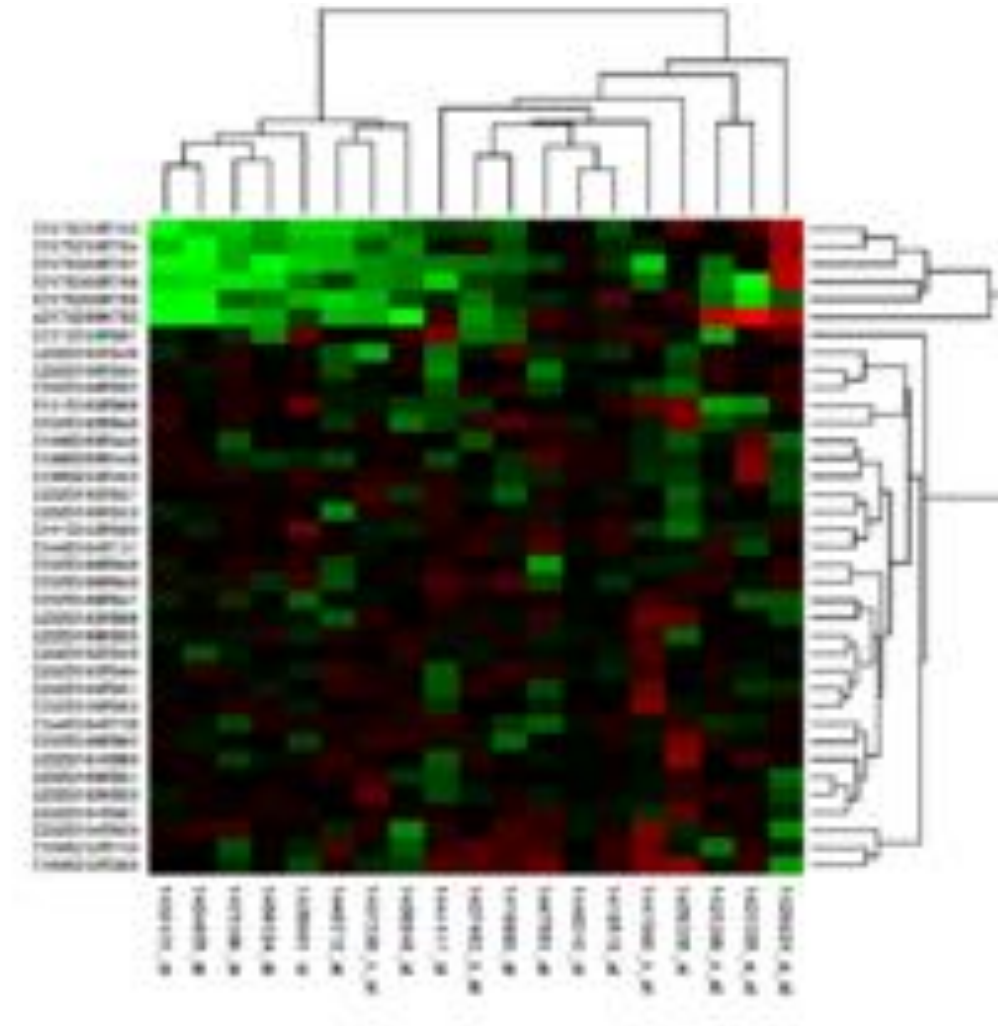
Commonly Used Visualizations



Heat map

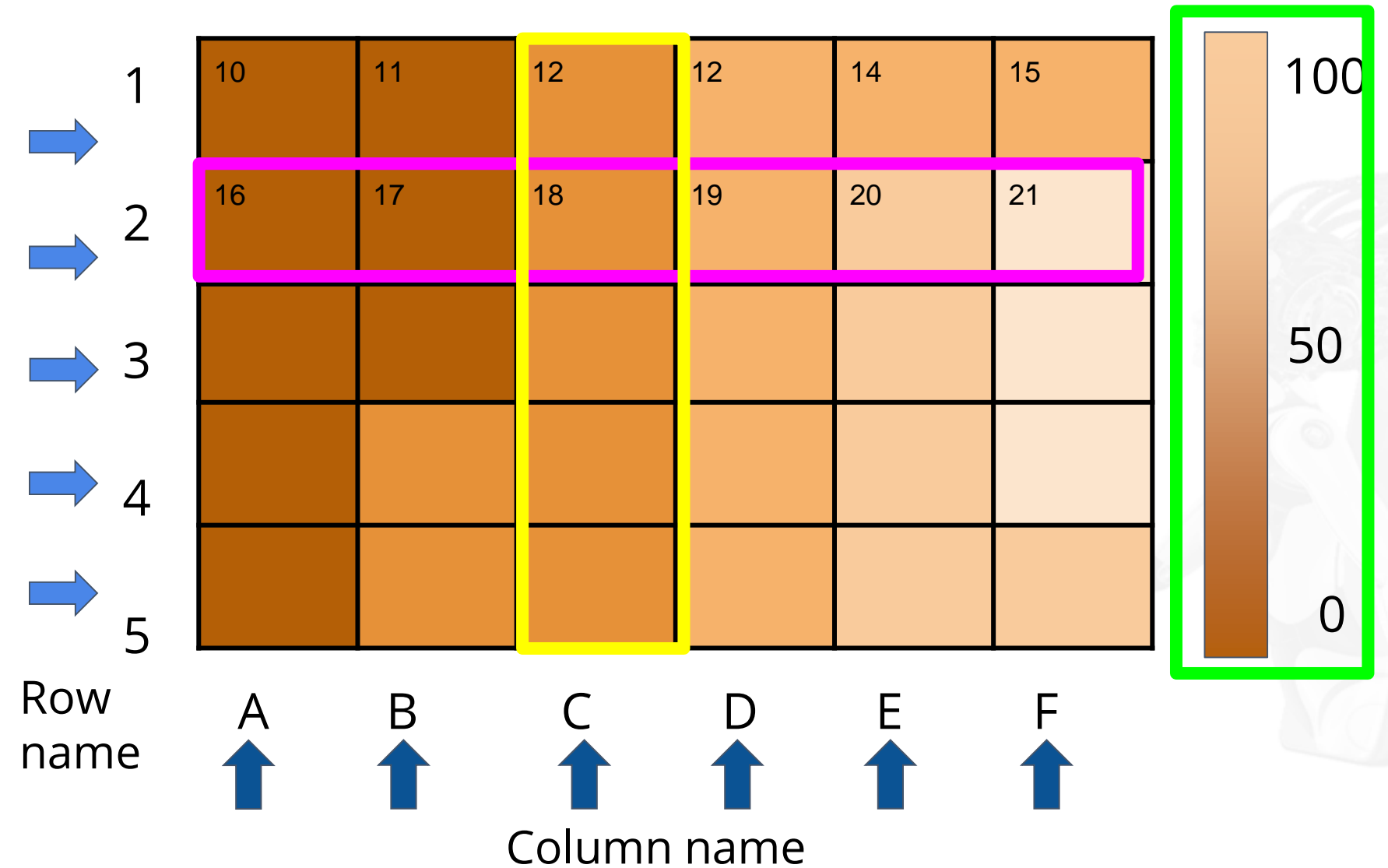
Heat map

- A heat map is a type of graph that uses a warm-to-cool color spectrum to visualize the data.
- It measures the relationship between multiple variables and shows the strength of relationships with colors.
- It helps in creating a visually impactful view of correlation.



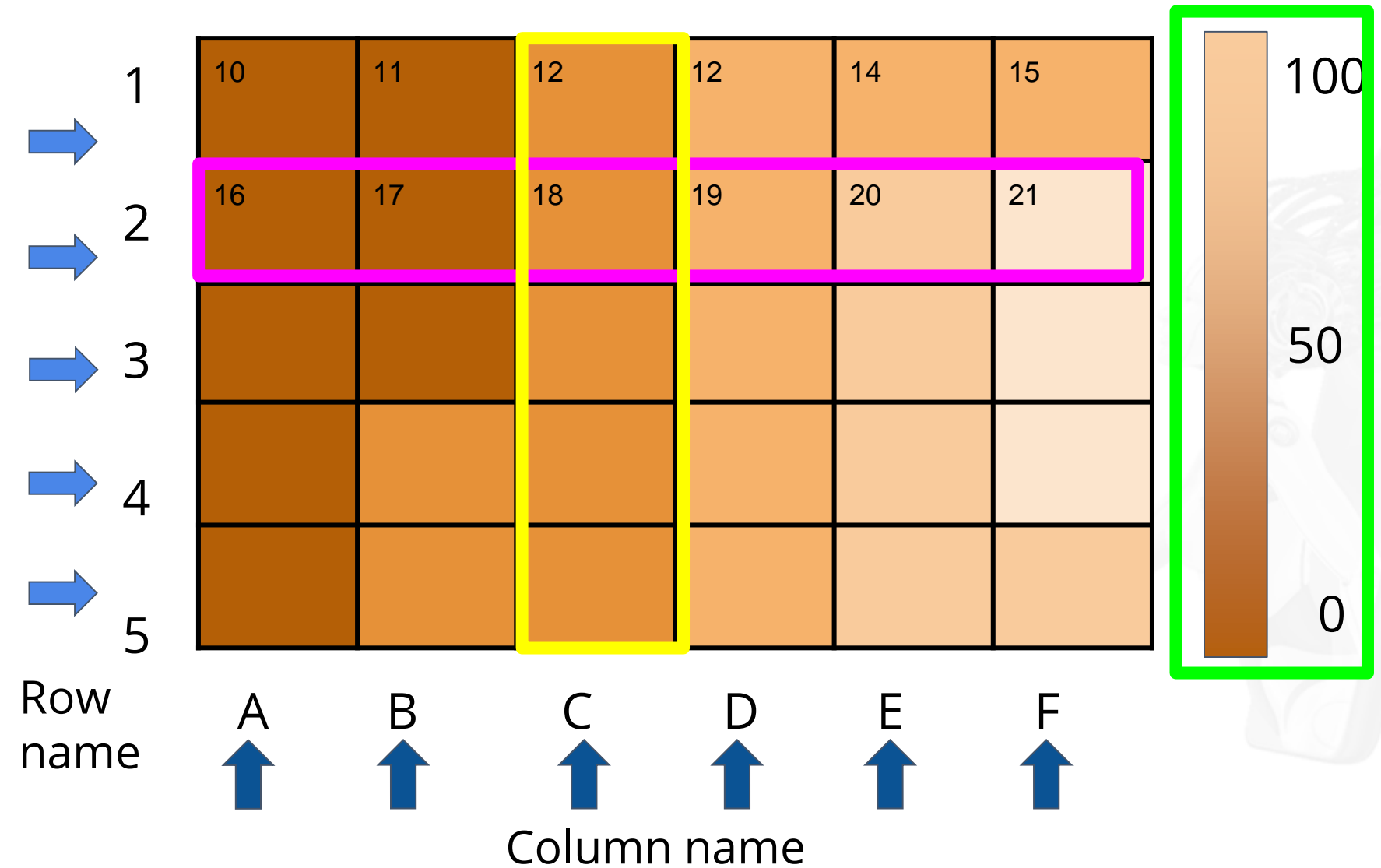
Heat map: Use Case

- All the rows are one category and all the columns are another category.
- Individual rows and columns are divided into subcategories.
- Cells either contain color-coded categorical data or numerical data.



Heat map: Use Case

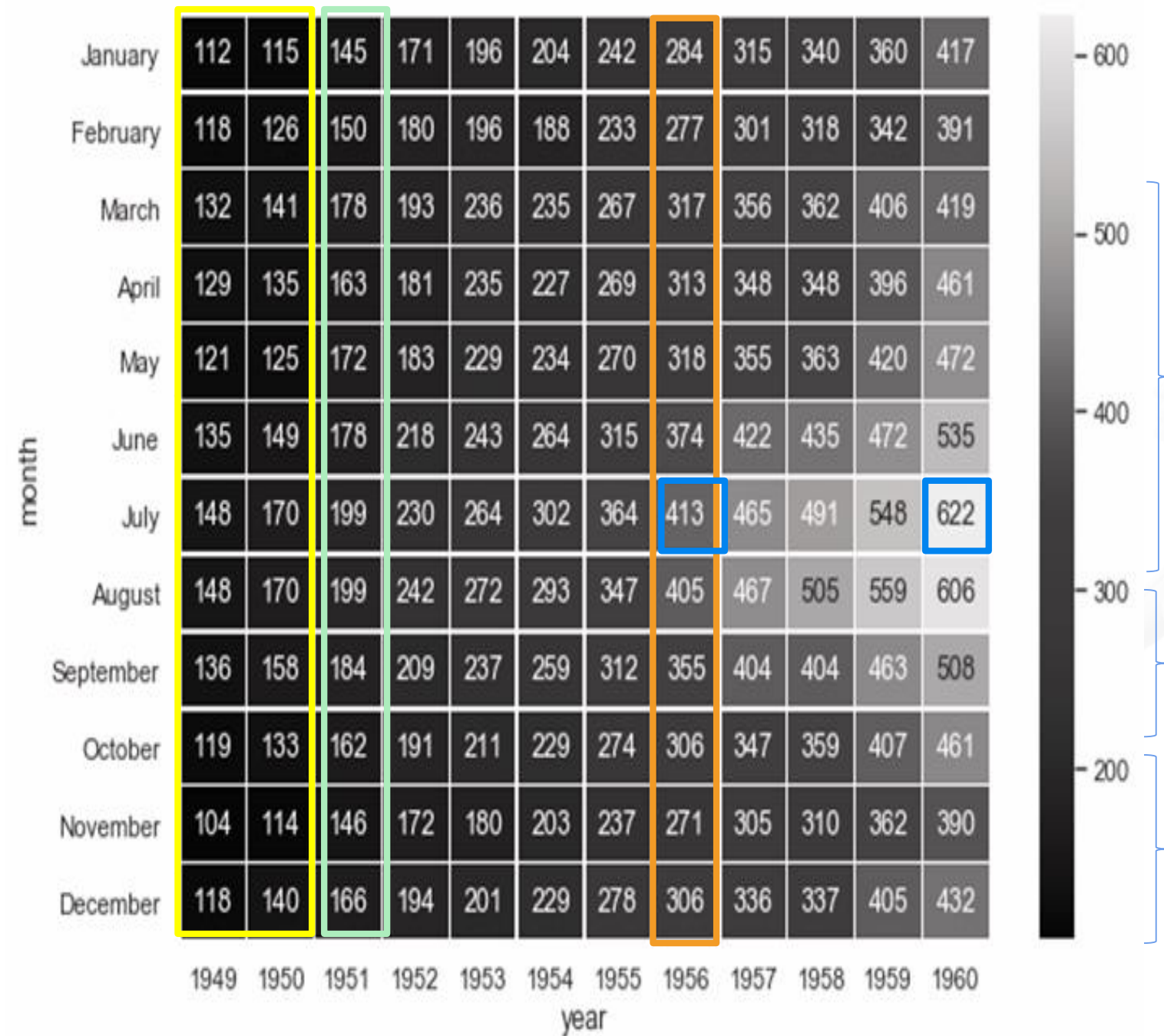
- Data in a cell is based on the relationship between two variables in the connecting row and column.
- Multiple value ranges can be represented by a selection of solid colors, while a single range can be shown by a gradient scale.



Heat map: Case Study

This heat map shows the sales data across months and year.

- It is observed that during 1949–1950, sales were in range of 0–200.
- Sales crossed the 200 mark in 1951 and increased every year.
- In 1956, sales were between 300 to 500 and the maximum sale was in July.
- Over the years, July had the maximum sales and it peaked in 1960.

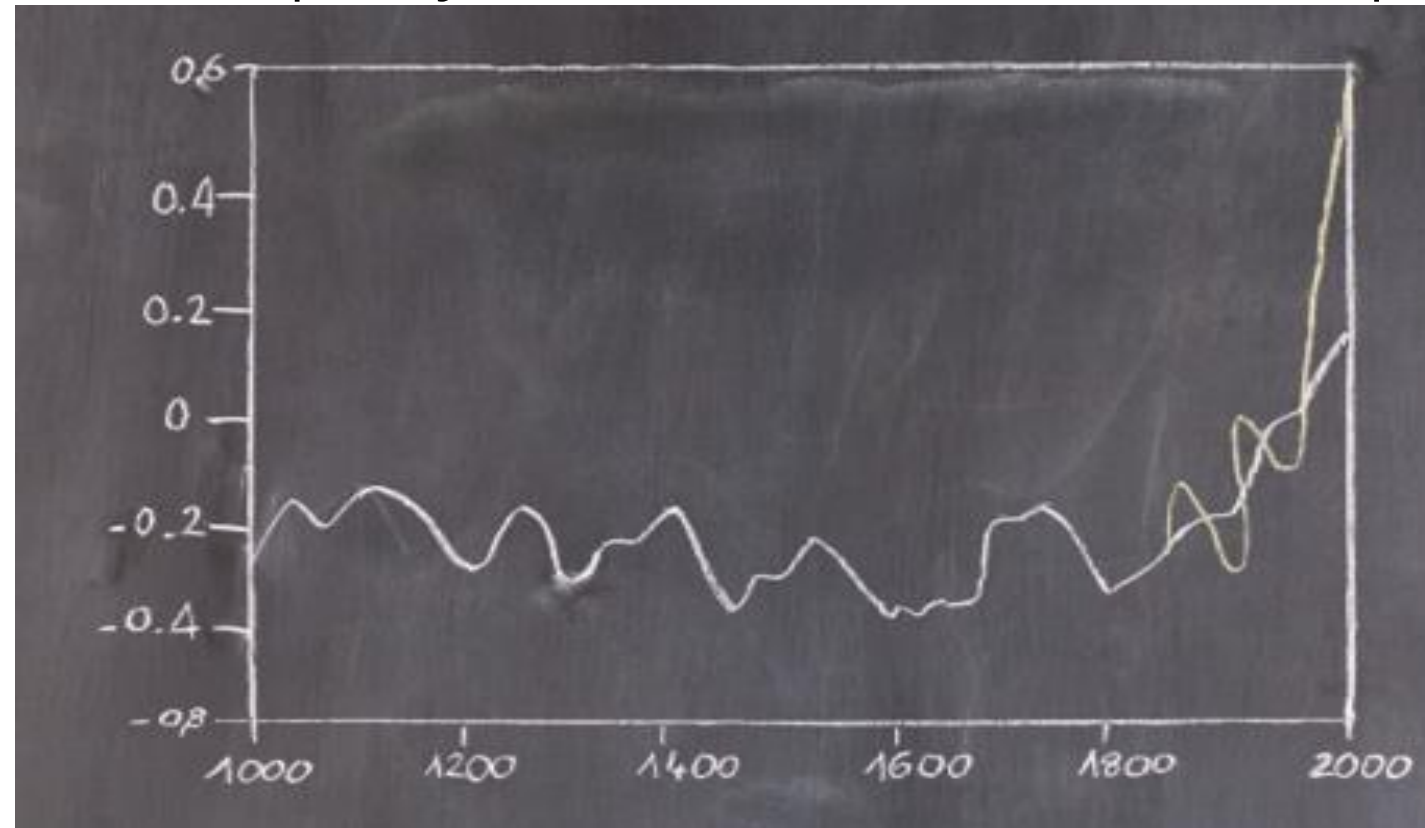


Frequency Distribution Plot

Frequency Distribution Plot

Frequency distribution plot measures the frequency of occurrence for a given value or range.

A normalized frequency distribution normalizes total frequency to one.



The frequency of an event is the number of times the event occurs in an observation.

Observations within a given interval are in graphical or tabular format.

Frequency Distribution Plot

Frequency distributions can be displayed in these formats:



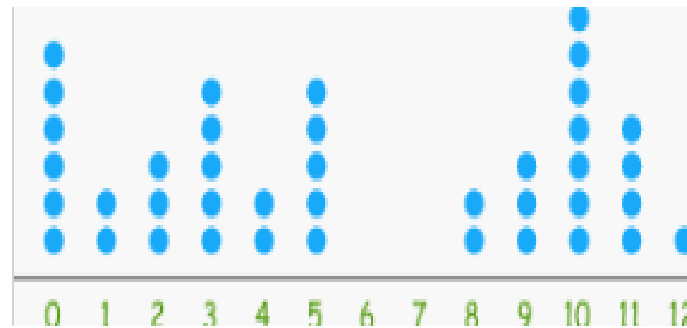
Table



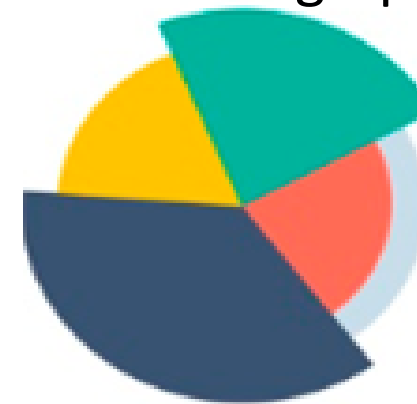
Histogram



Line graph



Dot plot

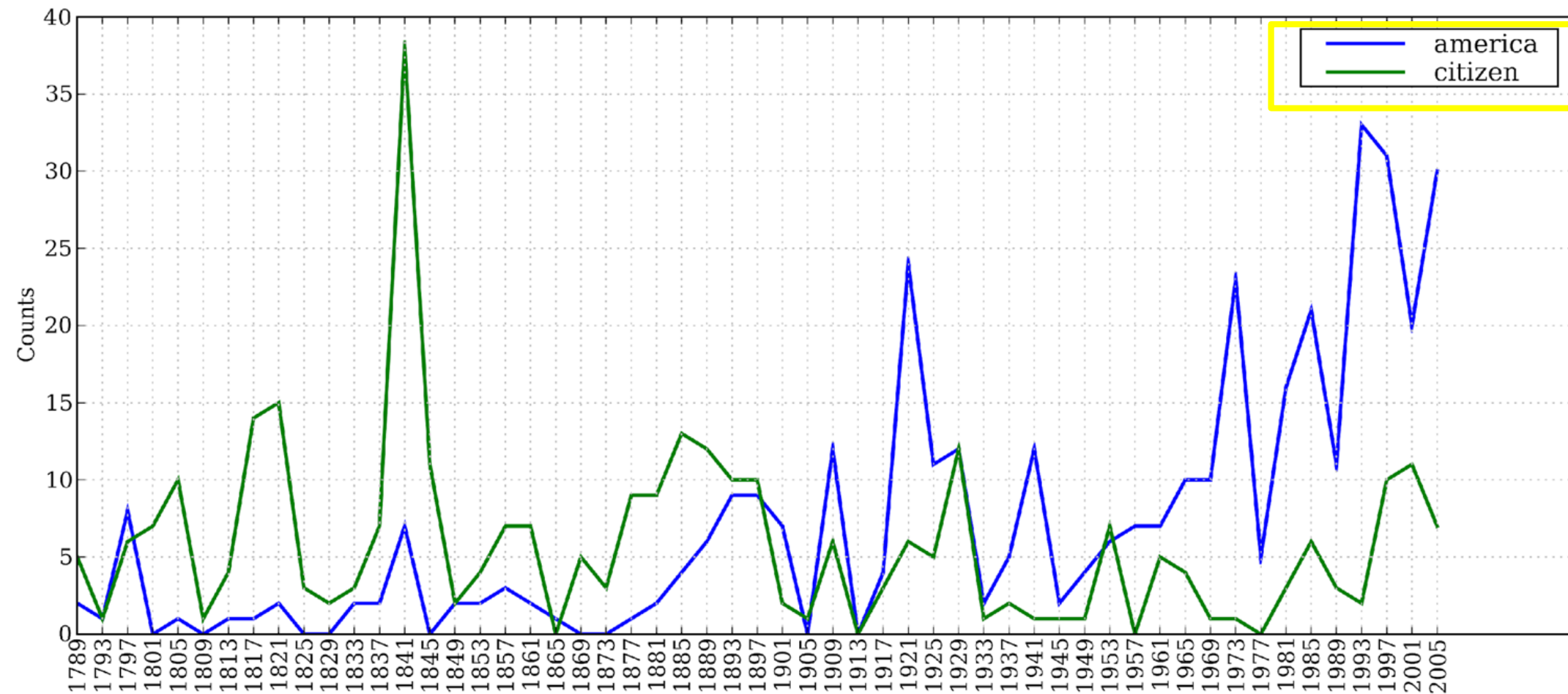


Pie chart

Analysts use the frequency distribution plot to check or illustrate the data collected in a sample.

Frequency Distribution Plot: Use Case

This conditional frequency distribution graph shows the usage of two specific words in public speeches over a period of time.

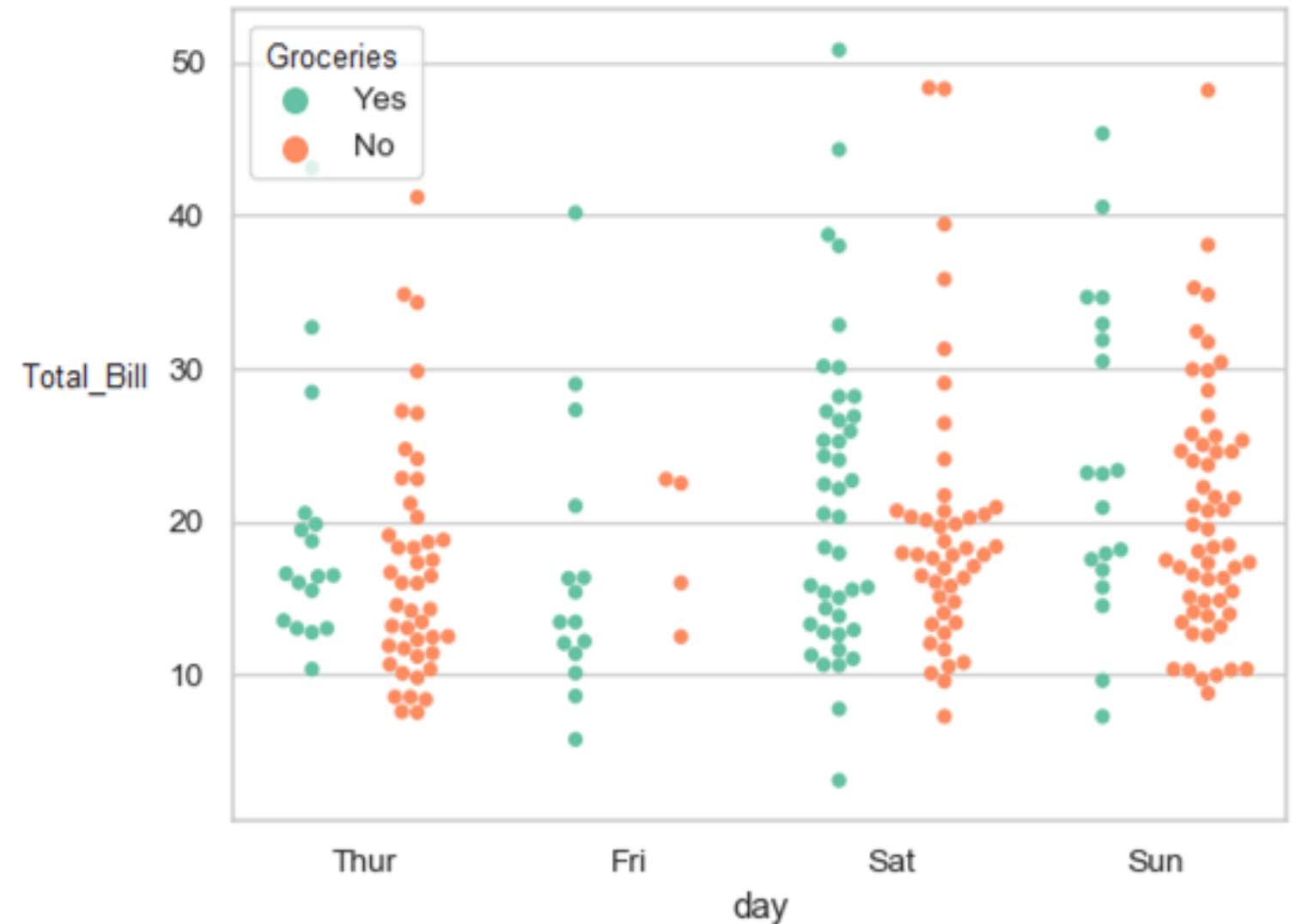


The graph indicates that the use of *America* has increased, while the use of *citizen* has reduced gradually in speeches.

Swarm Plot

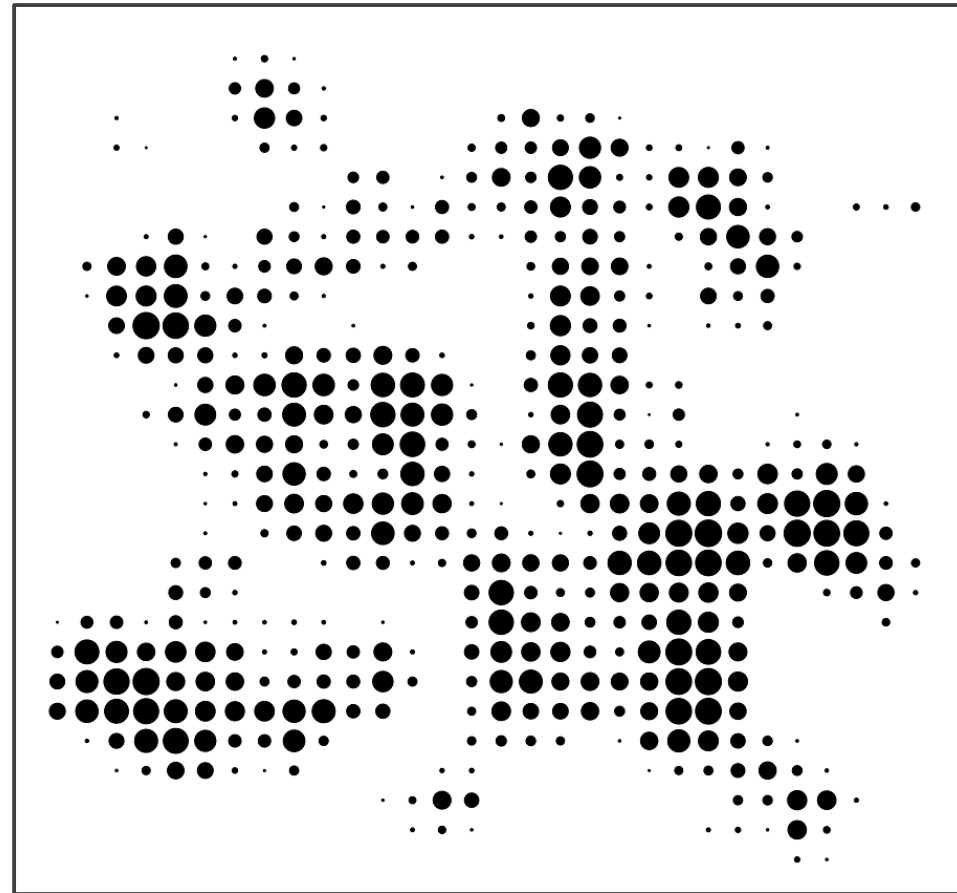
Swarm Plot

- A swarm plot gives a good representation of the distributions but works well only for small data sets.
- It is useful to examine individuals, places, or things in your data.
- It allows you to plot all of your points in a single space.
- It is a one-dimensional scatter plot as it plots the data on a single axis and then offsets in the other direction to show volume.



Swarm Plot

A swarm plot enables you to separate all overlapping points, making each point visible.



Beeswarm plot

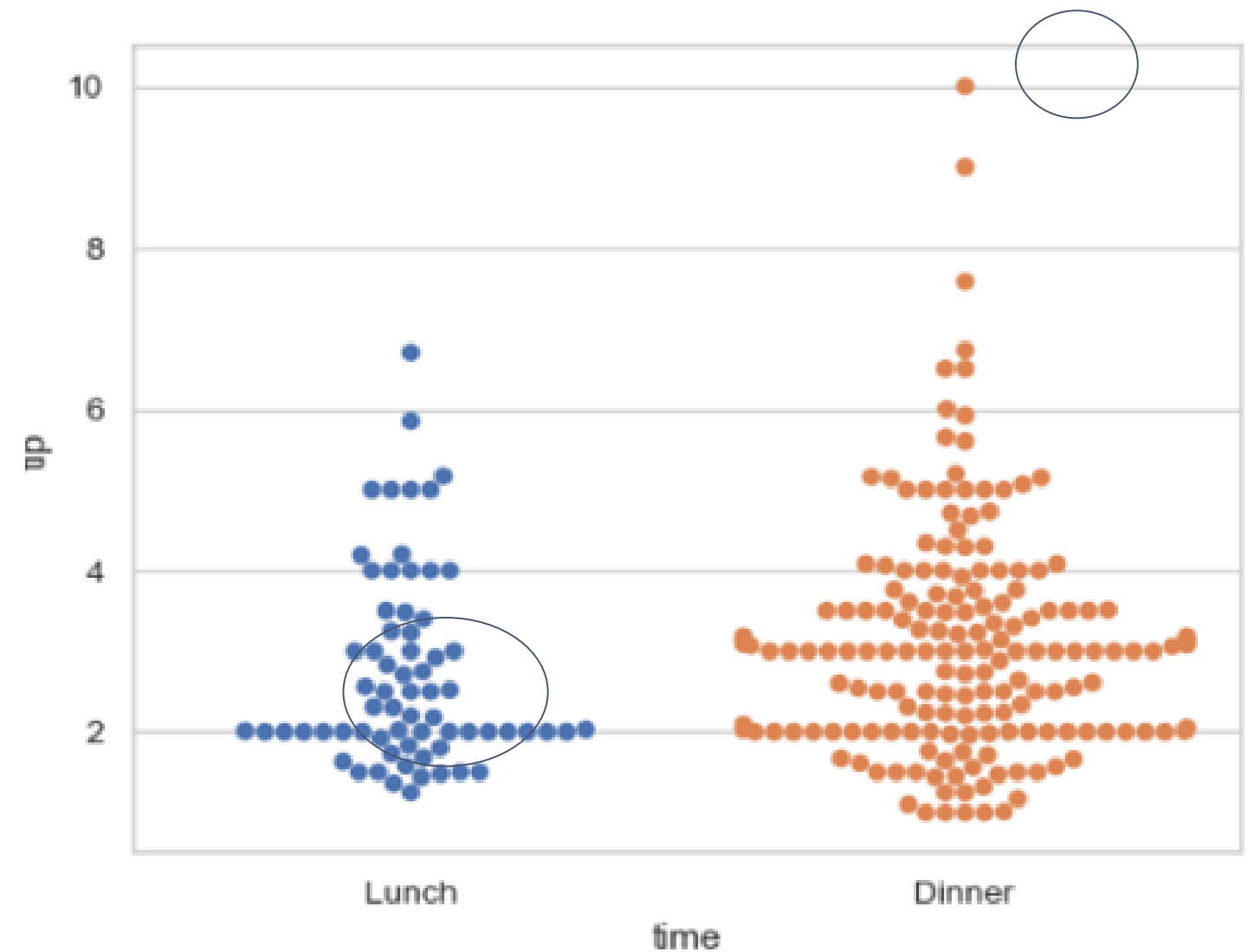
It is also called a beeswarm plot as the graphical representation is similar to a group of bees.



Swarm Plot: Use Case

In this swarm plot:

- The X-axis denotes time and the Y-axis denotes the tip amount
- The blue graph represents lunch and the orange graph represents dinner
- Tips are higher during dinner and most tips are \$2 and \$3
- The maximum tip is \$10, which was given for dinner.



Importance of Data Visualization

Importance of Data Visualization in Analytics

- Data visualization tools provide access to trends, outliers, and patterns in data.
- They help organize and present important findings from the data.
- Data in user-friendly charts help businesses gain insights to make right decisions.



Importance of Data Visualization in Analytics

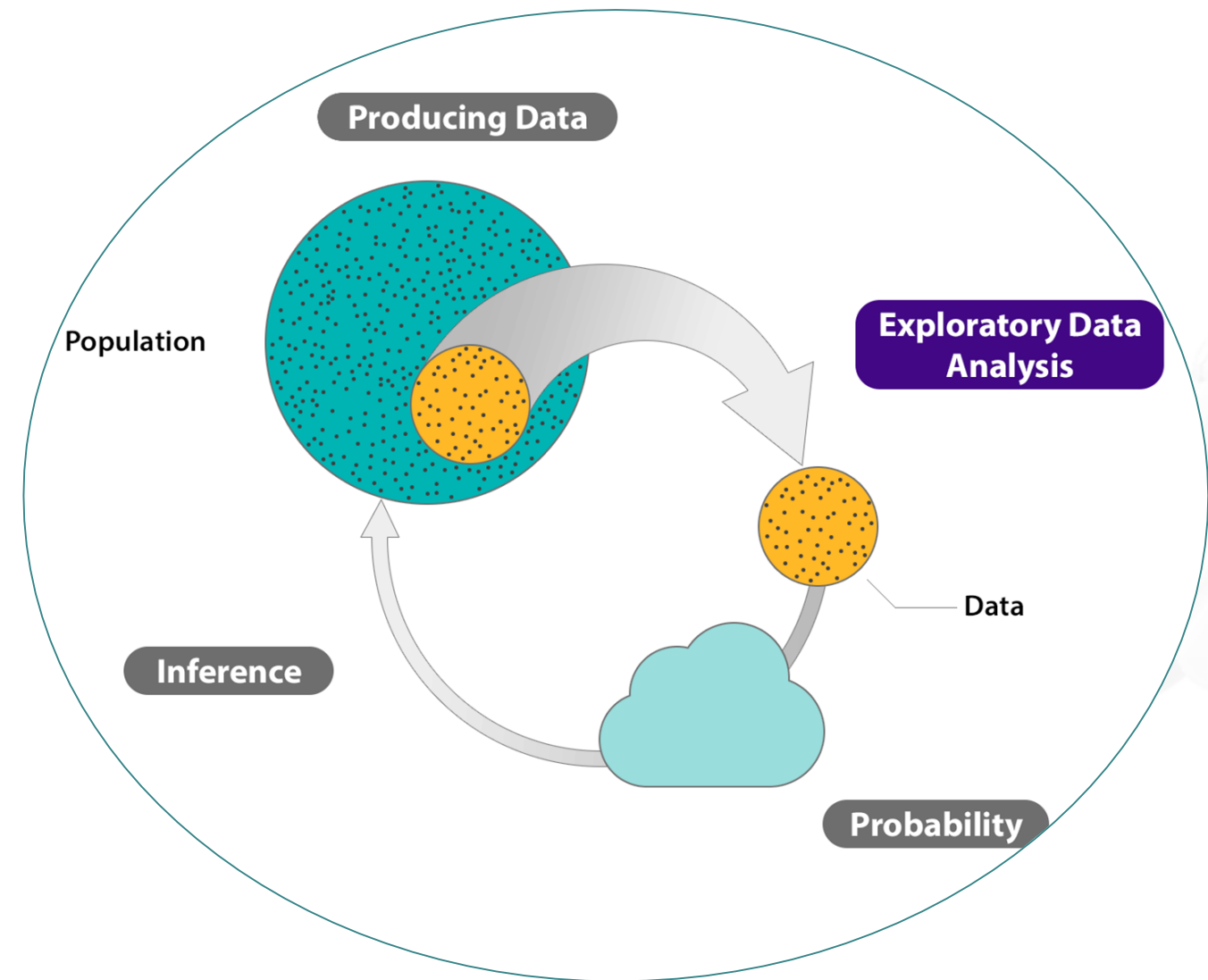


- Data analytics tool allows a user to present massive data intuitively.
- Decision makers see patterns, trends, and correlations in the data being analyzed.
- It help decision makers in cutting costs or improving operational processes.

Exploratory Data Analytics

Exploratory Data Analytics

- Exploratory data analytics is an approach to analyze data sets to summarize their main characteristics.
- Data visualization in exploratory data analytics is the first step towards modeling.
- EDA primarily helps analyze data beyond the formal modeling.



Exploratory Data Analytics



The steps involved in EDA are:

- Get detailed insights into the dataset
- Understand critical impact variables that influence the dataset
- Detect if any outliers are present in the dataset
- Test the underlying assumptions of the dataset

Data Visualization Tools

Data Visualization Tools

- Tableau is the most widely used data visualization tool due to its simplicity and ability to produce interactive visualizations.
- It has a large customer base of more than 50,000 accounts across many industries.



- FusionCharts is a JavaScript-based visualization package that can produce and integrate 90 different chart types.
- It has a range of *live* example templates so that you can simply plug in your data sources as required.



Data Visualization Tools

- Highcharts is often chosen when a fast and flexible solution has to be rolled out.
- Its cross-browser support feature helps users view and run interactive visualizations.



- Datawrapper has a simple interface that makes it easy to upload CSV data, create charts, and maps.
- It is becoming popular among media organizations to create charts and present statistics.

Datawrapper

Data Visualization Tools

- Plotly enables complex and sophisticated visualizations.
- It is integrated with analytics-oriented programming languages such as Python, R, and Matlab.



- Sisense provides a full stack analytics platform and simple-to-use drag and drop interface.
- It helps in creating charts and complex graphics with minimum hassle and provides a repository for gathering multiple sources of data.



Other Visualization Tools

Other Visualization Tools

- Power BI is a powerful suite of business analytics tools and has intuitive UI for users familiar with Microsoft products.
- It can create customized, user-defined visualizations as well as sophisticated 3D maps.



Power BI

- The Looker BI tool provides extensive visualization abilities, along with real-time analysis.
- Users can either use templates from the Looker library or create a custom visualization.



Other Visualization Tools

- Domo is self-service business intelligence that focuses on social collaboration.
- It provides real-time data and uses creative data displays such as multi-part widgets and sparklines.



- Board is a full-featured business intelligence system.
- It serves midsize and enterprise-level companies in different industry segments.



Other Visualization Tools

- Qlik Sense has a clean and clutter-free user interface and a highly customizable setup.
- Qlik Sense is Tableau's biggest competitor.
- It has over 40,000 customer accounts in more than 100 countries.



Languages and Libraries for Data Visualization

Languages for Data Visualization

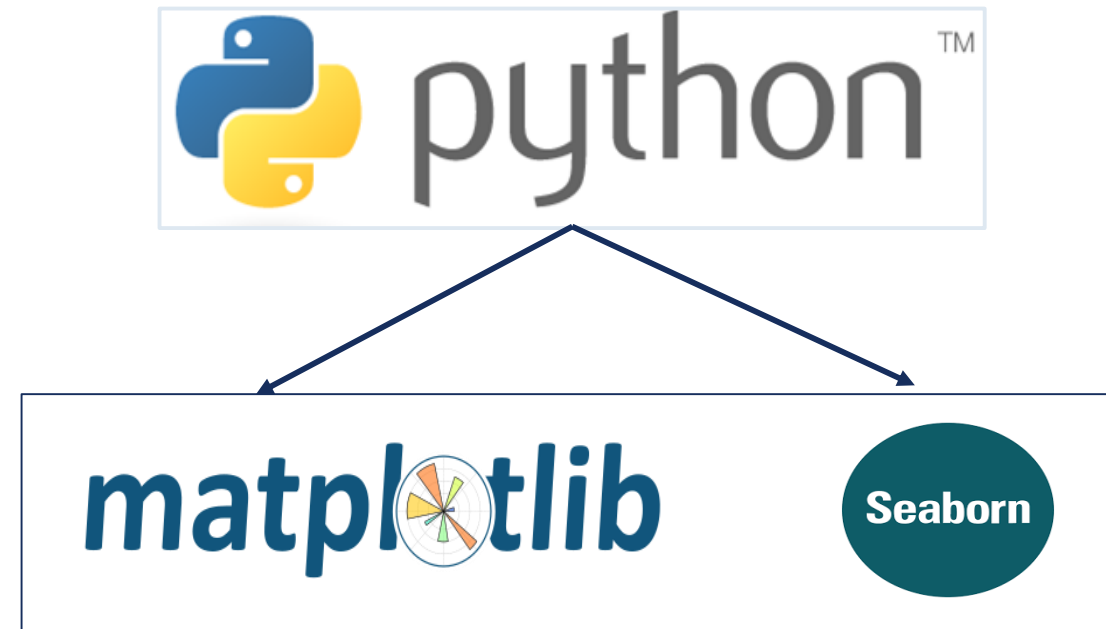
Few languages and libraries leveraged by data visualization:

- Scala
- R
- Python
- Javascript
- Java



Languages for Data Visualization

Python has two exclusive libraries for data visualizations that are Matplotlib and Seaborn.



Scala is a compiled language and the code written in Scala gets executed much faster.

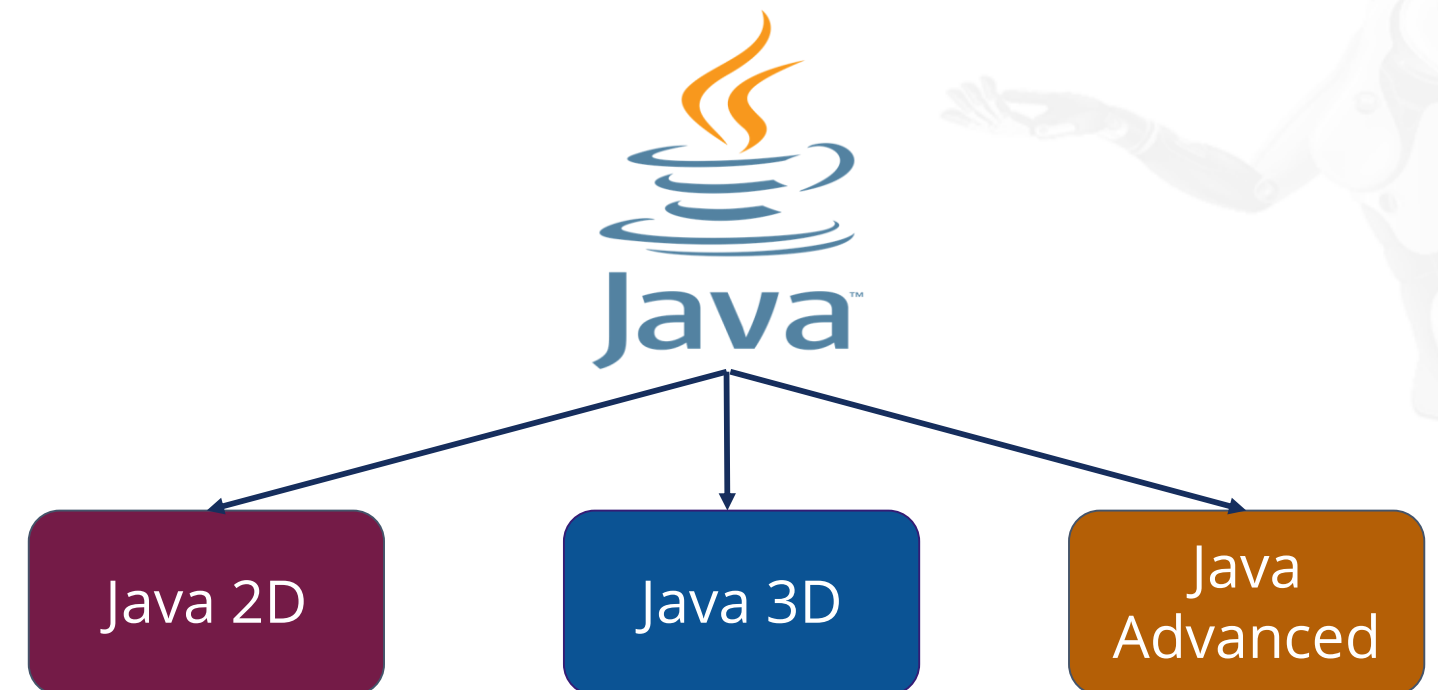


Languages for Data Visualization

Base graphics, lattice graphics, grid graphics, and ggplot2 are the four graphic systems supported by R.



Inbuilt libraries available in Java such as Java 2D, Java 3D, and Java advanced imaging makes data visualization simple with Java.



Data Visualization Libraries



Shiny is an R package and it is easy to build interactive web apps straight from R.

matplotlib

Matplotlib is the first data visualization library and has 2D and 3D graphics support.



Seaborn is a Python data visualization library and provides an interface for drawing statistical graphics.

bokeh

Bokeh is native of Python and helps to create interactive, web-ready plots by supporting streaming and real-time data.

Dashboard-Based Visualization

Dashboard-Based Visualization

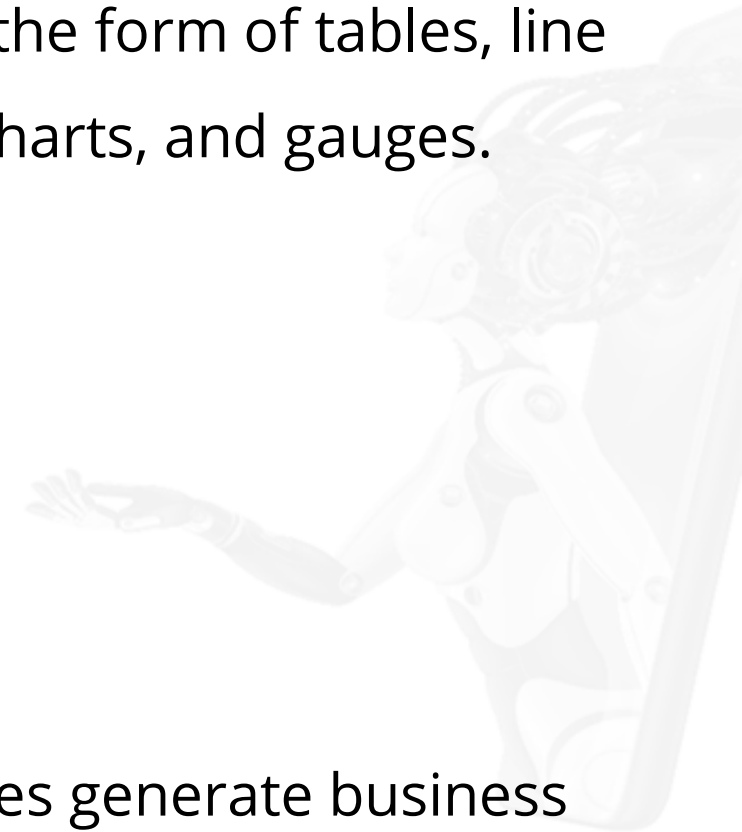
Provides a visual real-time representation of a company's data.



Monitors business health by visually tracking, analyzing, and displaying key data points.

Displays data in the form of tables, line charts, bar charts, and gauges.

Helps businesses generate business insights.



Dashboard-Based Visualization

This dashboard has parameters such as daily target, sales pattern, and other business insights from different charts.



Characteristics of effective dashboards:

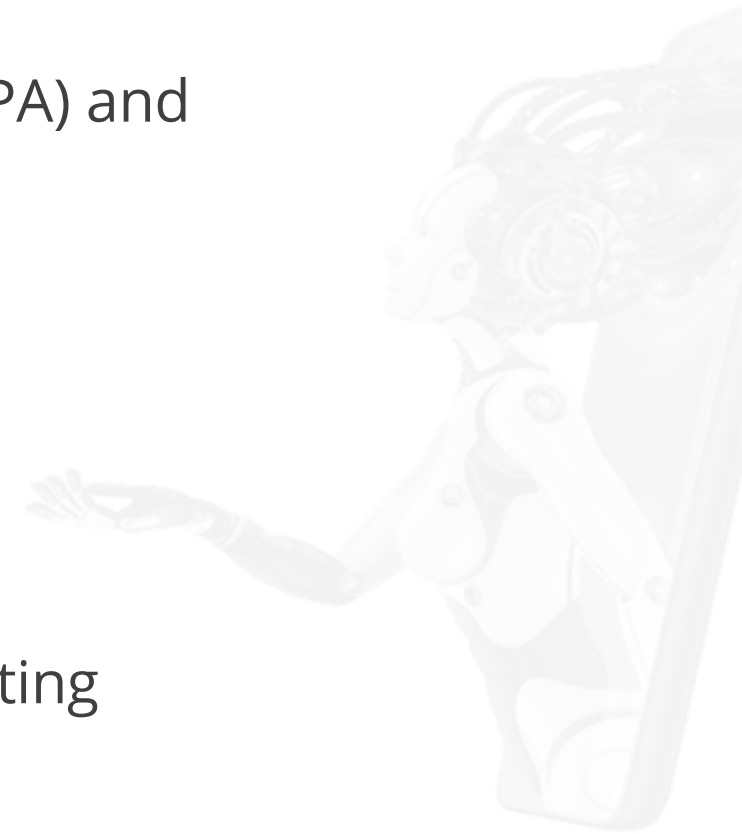
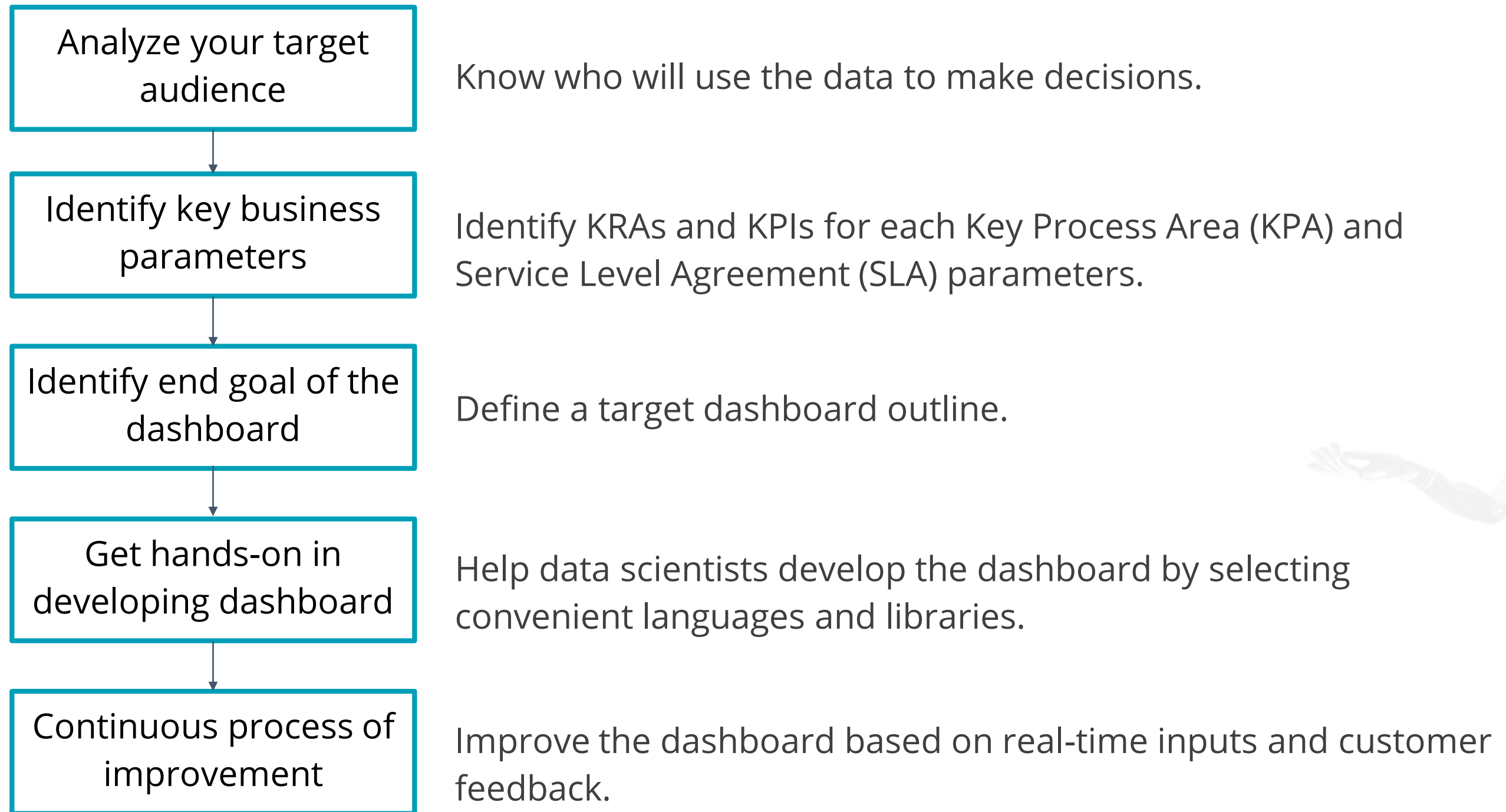
- Highly interactive
- Customizable interface
- Pulls real-time data from multiple sources

Steps for Dashboard-Based Visualization

- Analyze your target audience
- Identify Key business parameters
- Identify the end goal of the dashboard
- Get hands-on in developing the dashboard
- Continuous process of improvement



Steps for Dashboard-Based Visualization



BI and Visualization Trends

BI and Visualization Trends



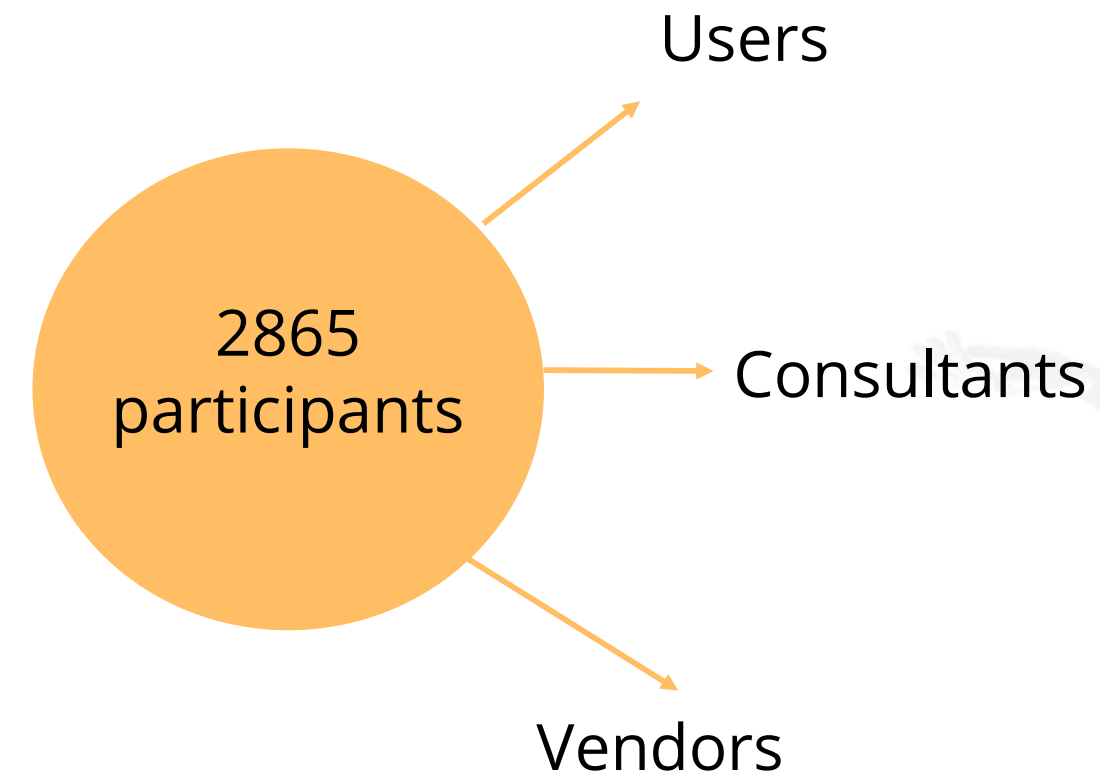
- The development of BI to analyze and extract value from various sources introduced many errors and low-quality reports.
- Companies choose to implement the Data Quality Management (DQM) policy as it is a key factor to efficient data analytics.

BARC Research Study

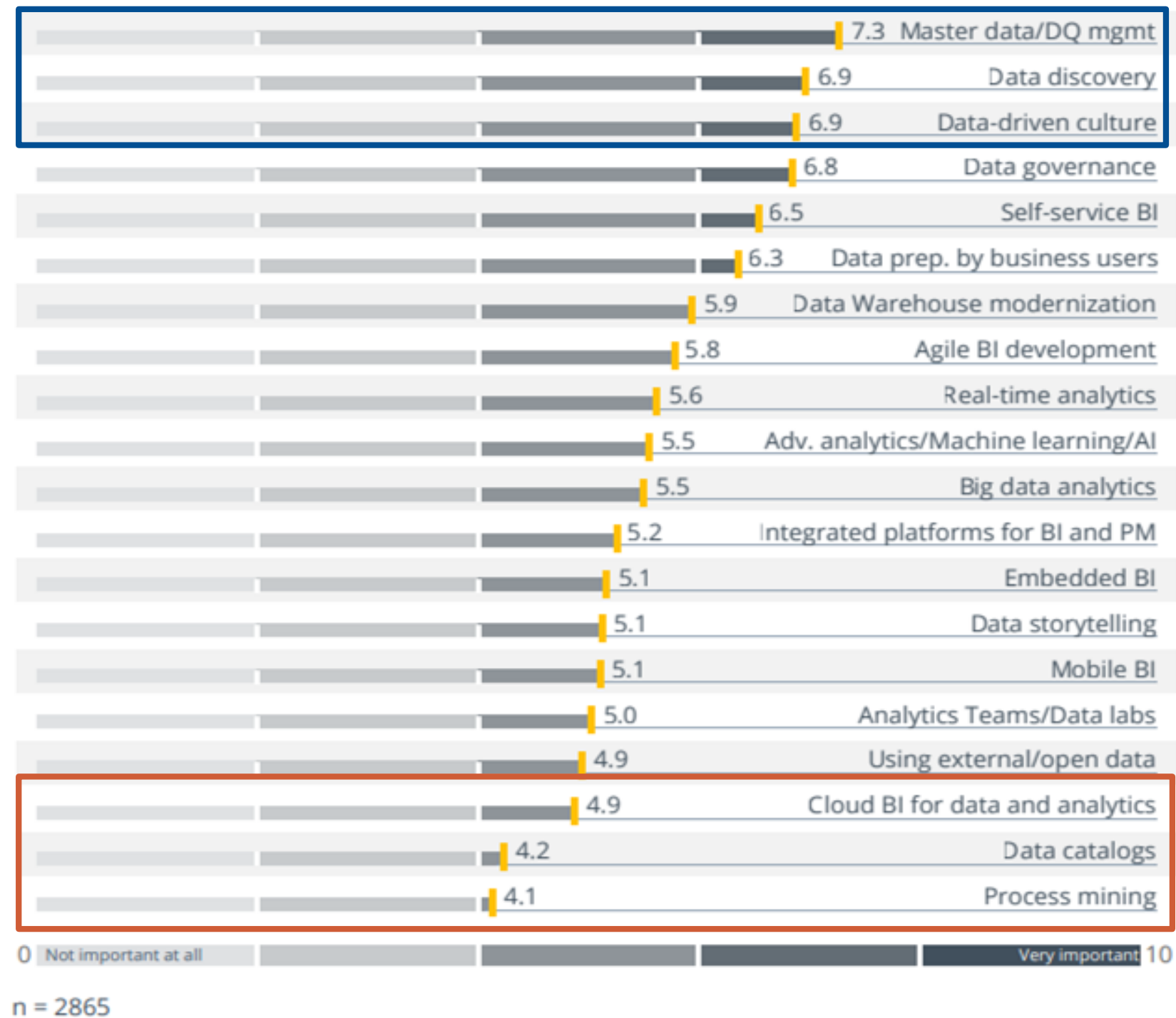
A survey was conducted by the Business Application Research Center (BARC) on the BI trends.

Users, consultants, and vendors were among the 2865 participants for the survey.

Master data/Data quality management was stated as the most important trend in 2020.

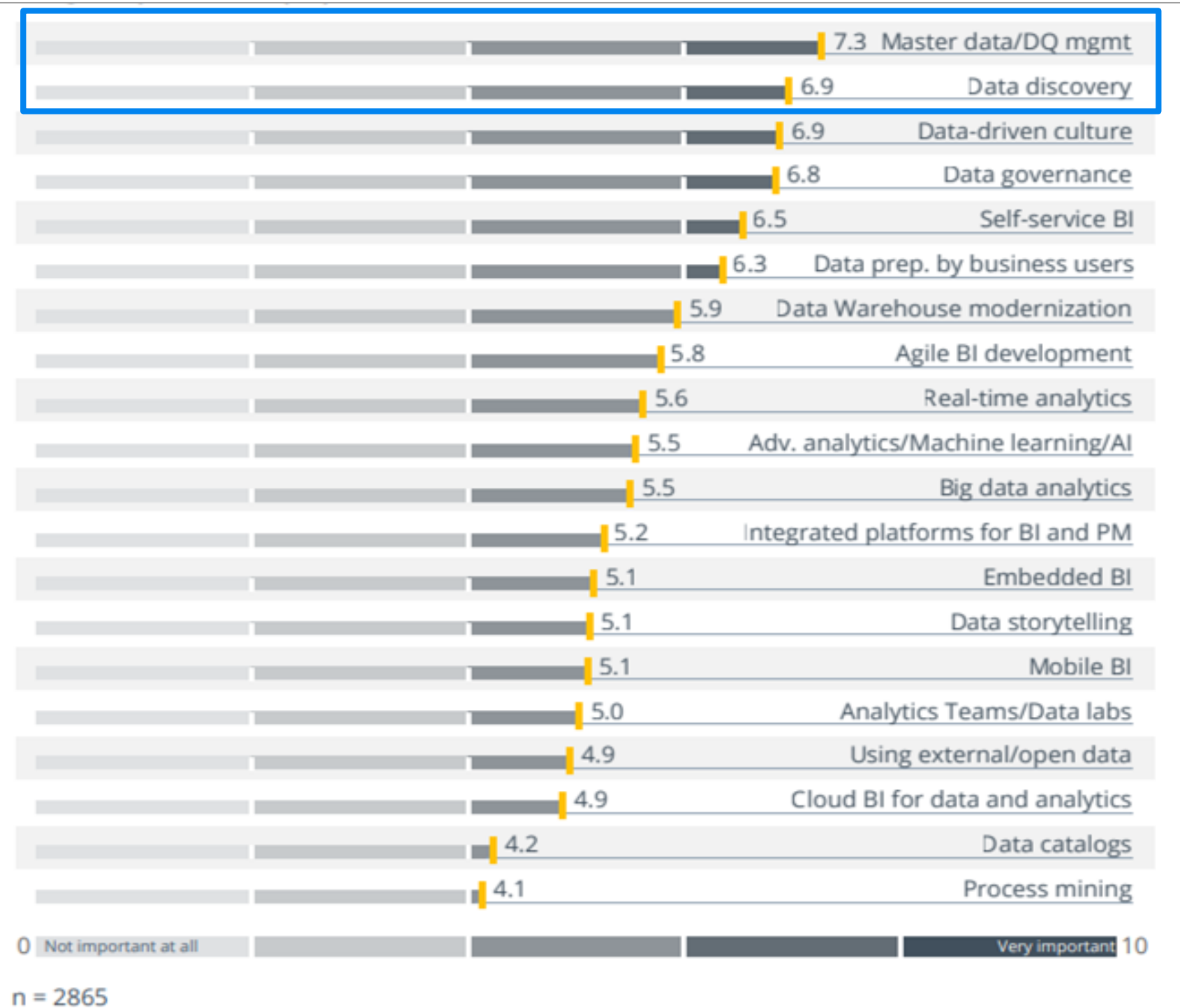


BARC Research Study



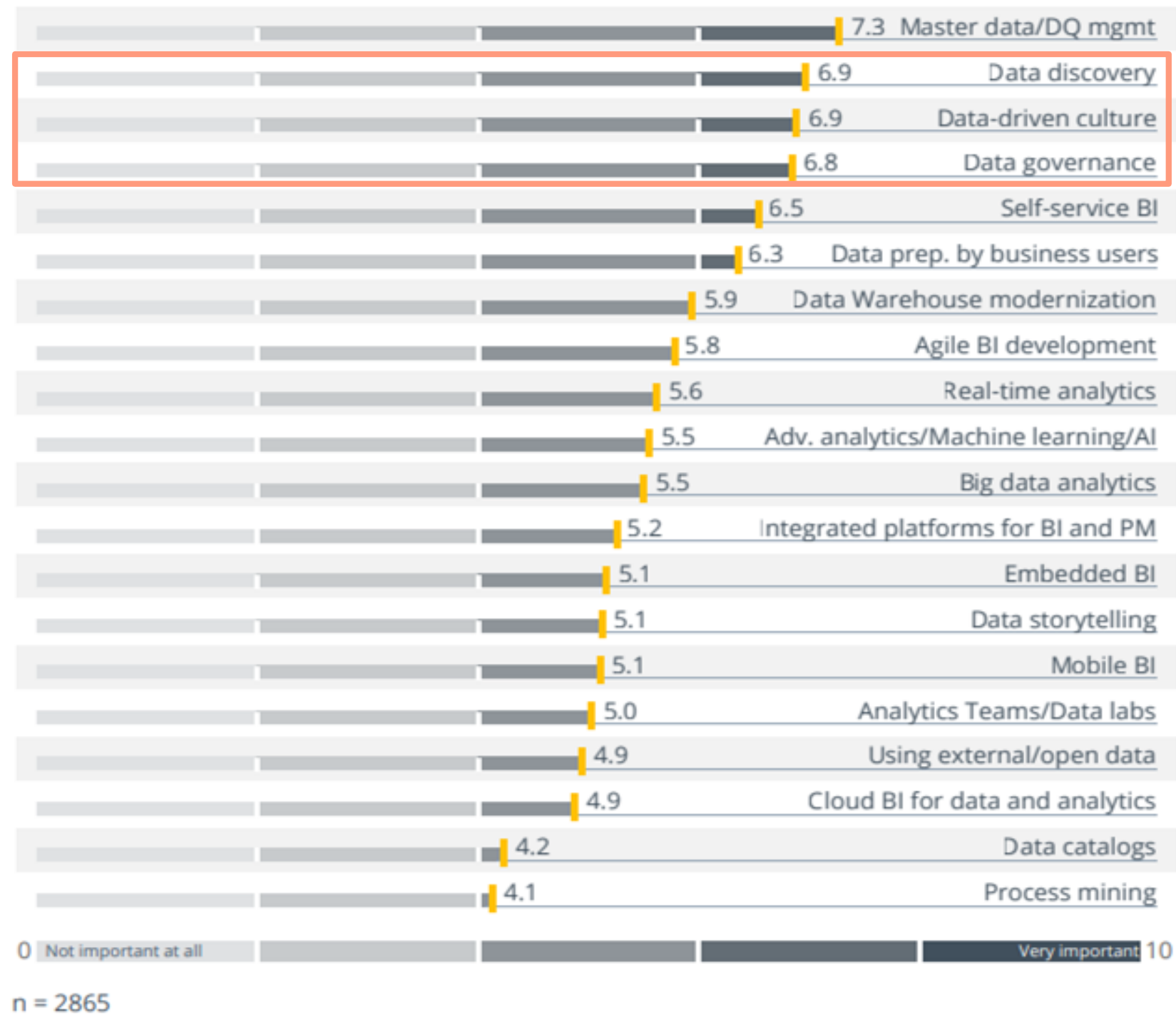
- BI practitioners identified *Master data/DQ management*, *Data discovery*, and *Data-driven culture* as the three most important trends in their work.
- *Cloud BI/data management*, *Data catalogs*, and *Process mining* were voted as the least important trends.

BARC Research Study



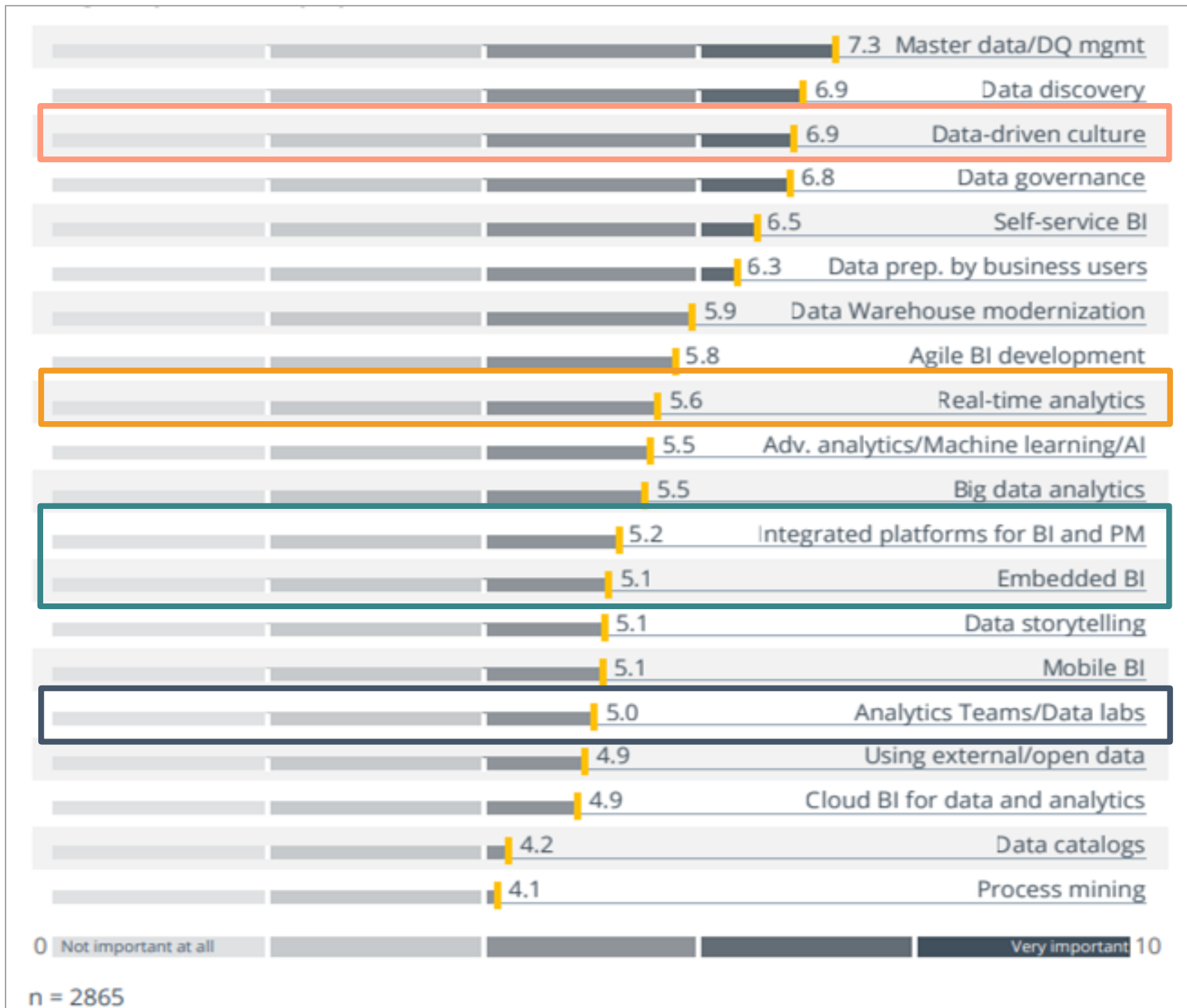
- *Master Data and Data Quality Management* build a strong foundation for handling data.
- *Data Discovery* describes how businesses collect data from various sources and then apply it to generate real business value.

BARC Research Study



- The significance of *Data discovery* shows a strong trend in business users empowerment.
- *Data-driven culture* depends on greater inclusion of various business departments.
- *Data governance* is an important trend due to the GDPR and increase in data security awareness.

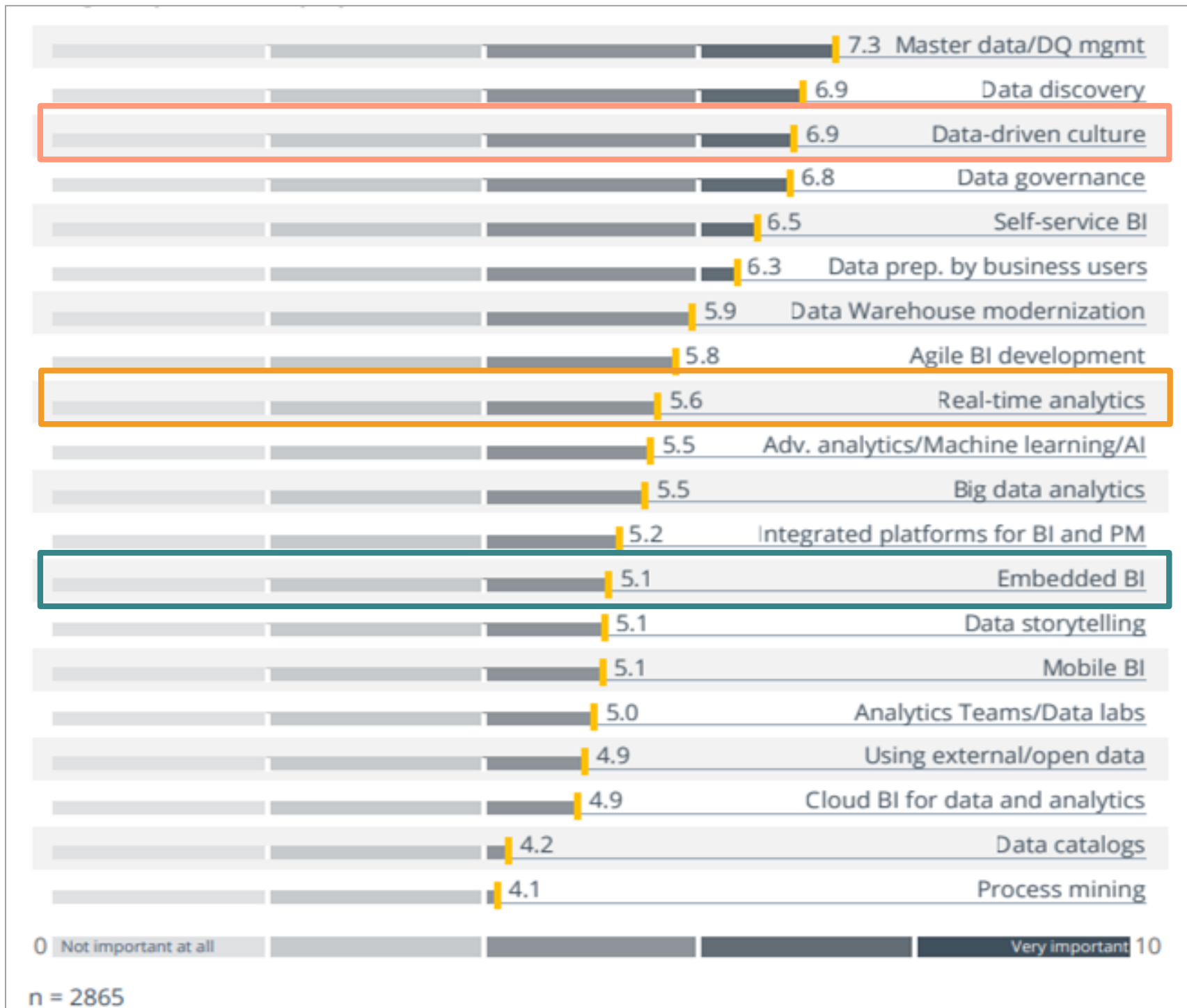
BARC Research Study



Trends that have increased in importance compared to the last year are:

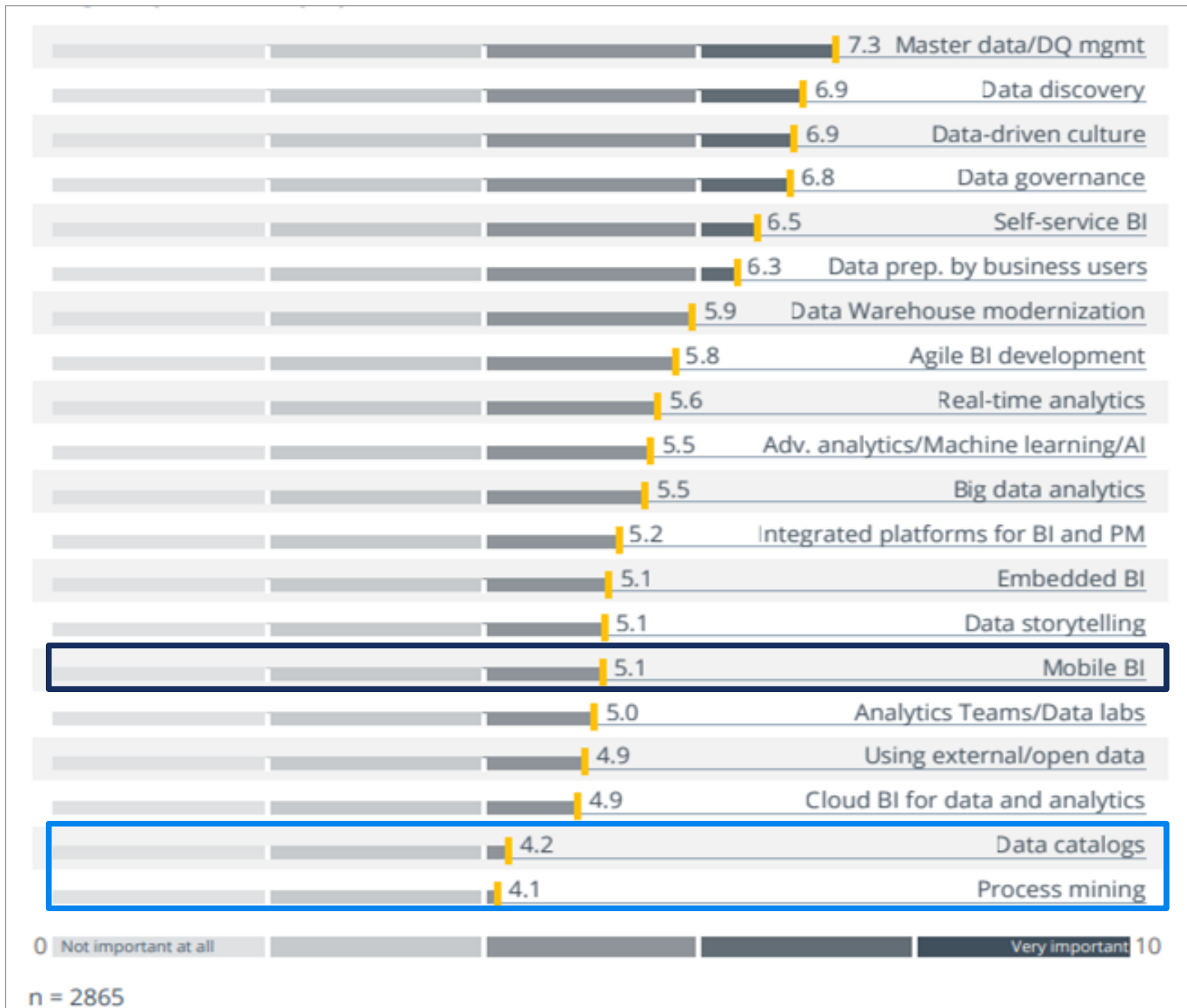
- Data-driven culture
- Real-time analytics
- Integrated platforms for BI and PM
- Embedded BI
- Analytics Team/Data labs

BARC Research Study



- In *Data-driven culture*, all decisions and processes are based on data and simple key figures like revenue.
- *Real-time analytics* is about capturing new data immediately after their occurrence and processing them for display or analysis.
- *Embedding intelligence* is growing its popularity in operational applications.

BARC Research Study



- *Mobile BI* has grown by only 20 percent in last eight years as the adoption was very slow.
- *Data catalogs* and *Process mining* trends are a little new in the market and has only recently been attracting more interest.

BI Software Challenges

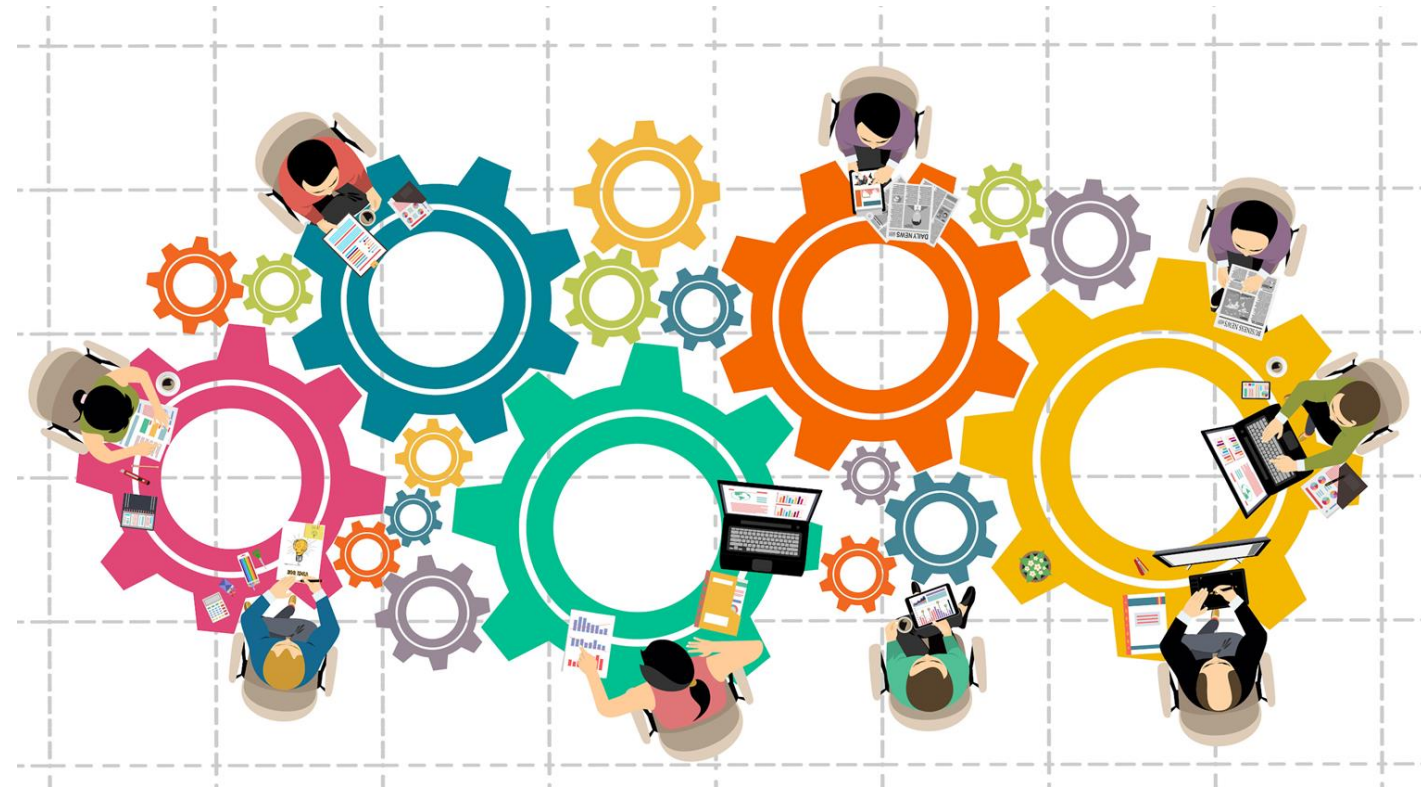
BI Software Challenges

- A survey by BI_Survey.com indicated that data quality is the top most problem for BI software users since 2002.
- BI software output is largely and highly impacted if the input data quality is not good.



BI Software Challenges

Poor quality input data leads to bad output results provided by any BI tool.



BI companies compile data in a usable system without affecting the validity of the original source.

Key Takeaways

- 🔗 Data visualization refers to the graphical representation of data using charts, graphs, and maps.
- 🔗 Heat map, frequency distribution plot, and swarm plot are commonly used visualizations.
- 🔗 Tableau, PowerBI, Datawrapper, and Sisense are some of the data visualization tools.

