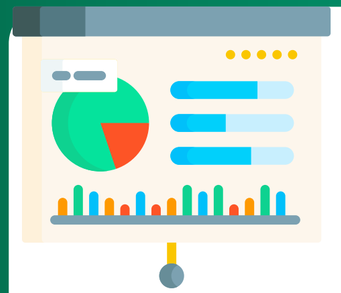
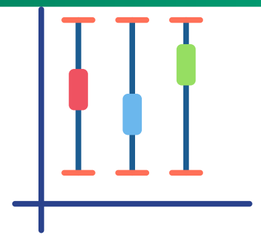
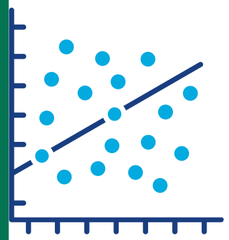
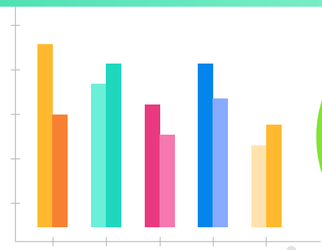
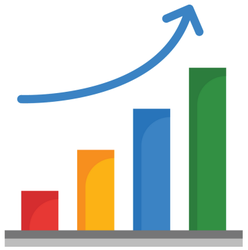
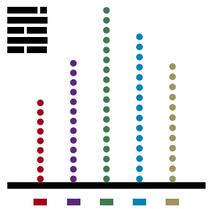





Day 2
Preparing Data for Analysis

Session 3

Graphical Presentation / Visualization







Session Outcome

After completing this session, researchers will be able to

- Summarize and Visualize Data
- Realize the Advantages of Visualization
- Realize the Importance of Visualization
- Visualize the Univariate, Bivariate and Multivariate Data
- Create Bar Charts, Pie Charts, Line Charts,
- Create Dot Plots, Box-Whisker Plots, Q-Q Plots etc.
- Interpret the Visual plots

Session Outline

- Introduction to Graphical Presentation or Visualization
- Objective, Importance, and Types of Graphical Presentation.
- Graphical Presentation for Different Datasets.
- Interpretation of Bar Charts, Pie Charts, Line Charts
- Dot Plots, Box-Whisker Plots. Q-Q plots.

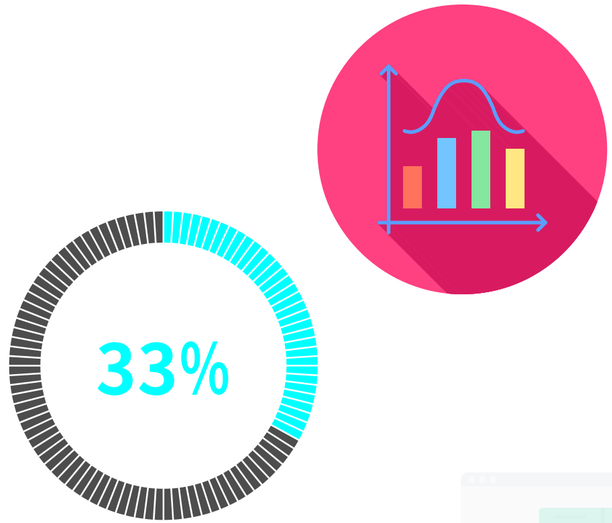
Different Summarization Techniques

Different Types of Summarization Techniques (Already described in the previous day)

Summarizing data in Exploratory Data Analysis (EDA) involves examining and describing the key characteristics, patterns, and relationships within a dataset to gain insights and inform the further analysis. There are two main methods for summarizing data:

- Summarization in Tabular Form

1. Frequency Distribution
2. Cumulative Distribution
3. Relative Frequency Distribution
4. Percentage Distribution



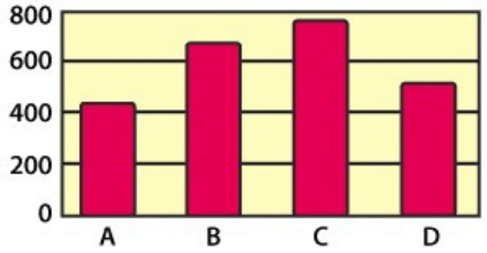
Different Summarization Techniques

- Visual Summarization

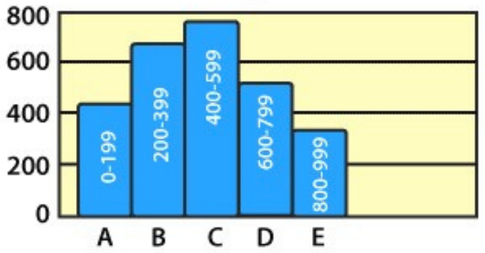
1. Bar graph
2. Pie chart
3. Line chart
4. Histogram
5. Dot plots
6. Box-whisker plots
7. Q-Q plots

Graphical Presentation

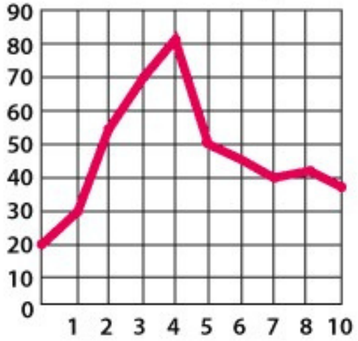
Bar Graphs



Histograms



Line Graphs



Stem and Leaf Plot

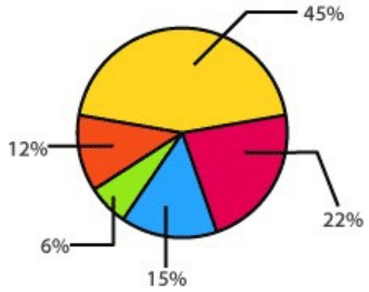
Stem	Leaf
0	1, 1, 2, 2, 3, 4, 4, 4, 4, 5, 8
1	0, 0, 0, 1, 1, 3, 7, 9
2	5, 5, 7, 7, 8, 8, 9, 9
3	0, 1, 1, 1, 2, 2, 2, 4, 5
4	0, 4, 8, 9
5	2, 6, 7, 7, 8
6	3, 6

Key : 6 | 3 = 63 Year

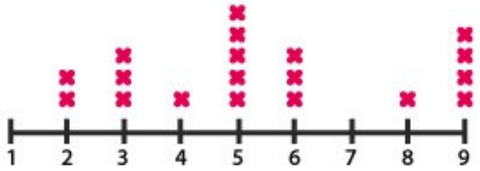
Frequency Table

Rulers of France		
Reign (Years)	Tally	Frequency
1-15	𝍸 𝍸 𝍸 III	18
16-30	𝍸 𝍸 I	11
31-45	𝍸 I	6
46-60	IIII	4
61-75	I	1

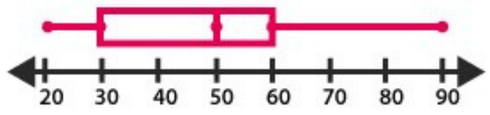
Circle Graph



Line Plot



Box and Whisker Plot



Graphical Presentation

Graphical presentation enhances communication, analysis, and decision-making by transforming complex data into visual formats, facilitating faster interpretation of patterns through graphs, charts, diagrams, and maps.

Objectives and Importance of Graphical Presentation:

Enhances Clarity and Understanding: Graphs, charts, and diagrams provide visual representations of data, aiding in understanding trends, patterns, and relationships.

Improves Communication of Findings: Visual representations effectively communicate key findings and insights..

Graphical Presentation

Makes Reports More Engaging: Colorful charts and graphs make reports more appealing and engaging for stakeholders

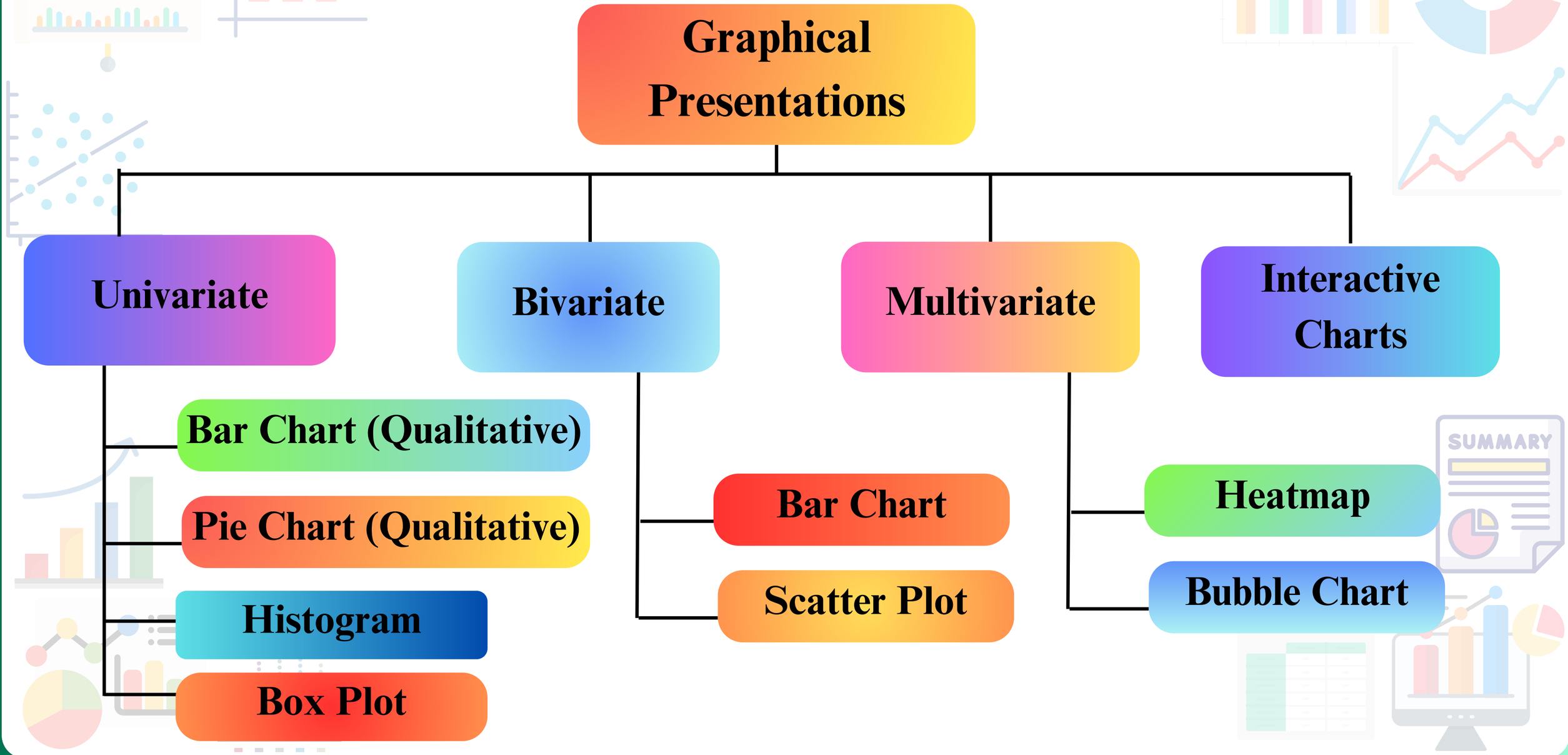
Empowers Decision Making: Visual representations empower educators, administrators, and policymakers to make informed decisions based on data.

Enables Comparative Analysis: Graphical presentation enables easy comparison of data sets, identifying disparities and areas of improvement.

Makes Complex Information Accessible: Graphs and charts make complex information more accessible to diverse audiences.

Strengthens Evidence-Based Reporting: Visual representations provide concrete evidence to support arguments, recommendations, or initiatives.

Types of Graphical Presentations



Graphs for Qualitative Data

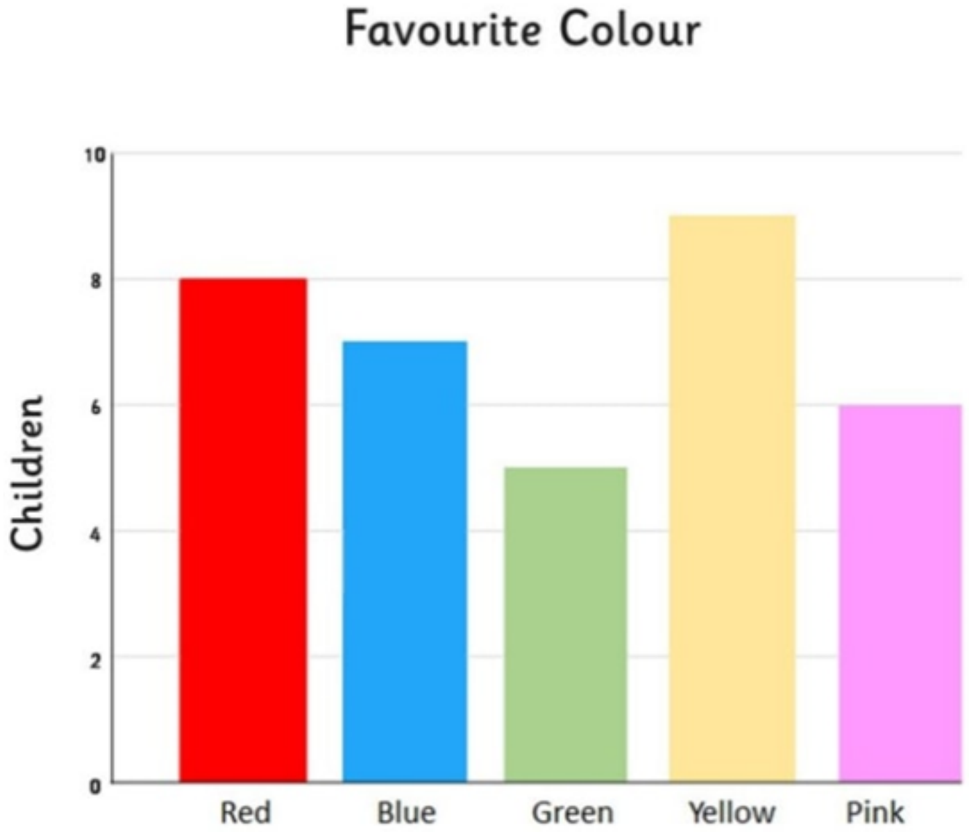
1. **Bar Diagram/ Chart**
2. **Pie Chart**

Bar Chart: A bar chart is a visual representation of categorical data, where bars of equal width are drawn to represent the frequency or proportion of each category. It's effective for comparing the values of different categories or groups.

- Mark categories on horizontal axis.
- All categories represented by intervals of same width.
- Mark frequency on vertical axis.
- Draw one bar for each category, with height representing frequency.
- Leave a small group between adjacent basis.

Graphs for Qualitative Data

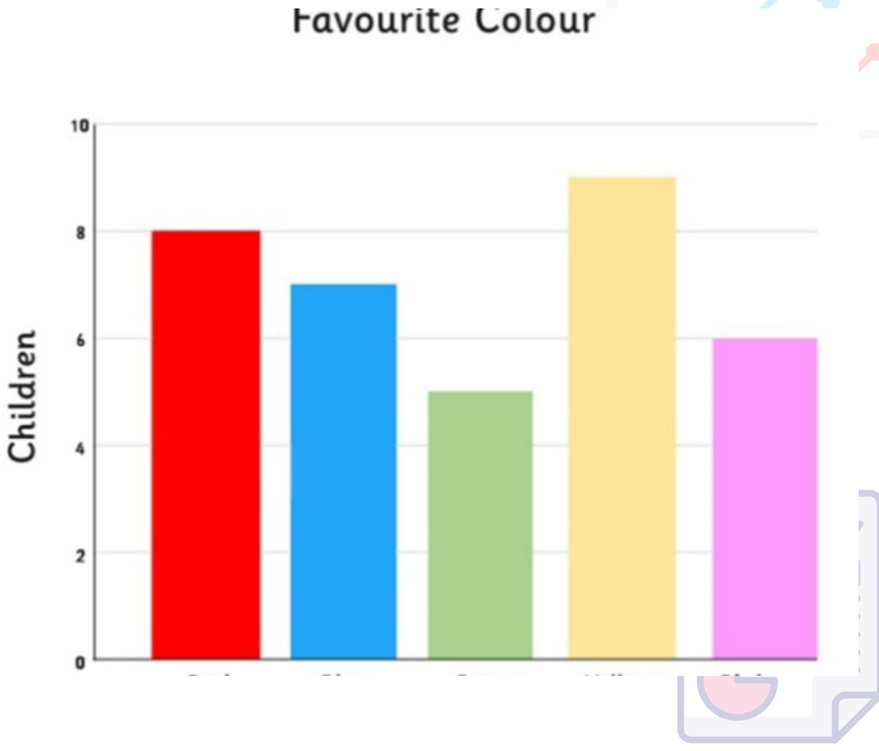
Example: Here is a vertical bar graph showing the popularity of different colours among a group of children.



Graphs for Qualitative Data

Interpretation

- Colors are represented by names and numbers labeled 'children'.
- Categories or categorical variables are written along the bottom.
- The numbers on the left represent the value of each category.
- To read a bar chart, consider the length of the bar connected to each category.
- Values are determined by the number that aligns with the bar for each color.
- The number that falls in a gap between two given numbers is the value.
- The popularity of a color grows as children age.



Graphs for Qualitative Data

Pie Chart: A chart that shows the proportion or percentage that each class represents of the total number of frequencies is called a pie chart.

- A pie chart is more commonly used to display percentages.
- The (whole pie) represents the total sample or population.
- We divide the pie into different portions that represent the different categories.
- To construct a pie chart, we multiply by the relative frequency of each category to obtain the degree figure or size of the angle for corresponding categories.

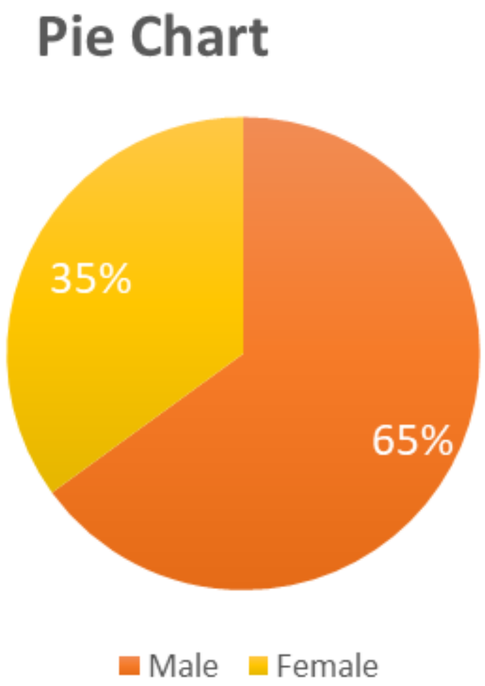
Example: Given a qualitative data, Gender of 20 students

F	M	M	M	F	M	M	F	M	M
M	M	F	M	M	F	F	M	F	M

The frequency distribution of Gender of 20 students are given bellow:

Graphs for Qualitative Data

Gender	Frequency	Relative Frequency	Percentage
Male	13	$\frac{13}{20} = 0.65$	$\frac{13}{20} \times 360^\circ = 234^\circ$
Female	7	$\frac{7}{20} = 0.35$	$\frac{7}{20} \times 360^\circ = 126^\circ$



Interpretation:

The pie chart illustrates the gender distribution among the sample, with males comprising approximately 65% (234°) and females around 35% (126°) of the total, indicating a majority of males in the sample population.

Univariate Visualizations

Univariate visualizations represent the distribution of a single variable, illustrating its frequency, range, and distribution characteristics without considering relationships with other variables.

Histograms: Histograms are graphical representations of the distribution of numerical data, displaying the frequency of values within predefined intervals or bins along the x-axis, providing insight into the data's shape, central tendency, and variability. The bars in a histogram are drawn adjacent to each other with no gap between them.

Histogram Drawing Process:

- Mark class boundaries on the horizontal axis.
- Draw frequencies or relative frequencies on the vertical axis.
- Draw a bar for each class, representing frequency.

Univariate Visualizations

Rules for constructing a class boundary:

1. Add 0.5 with the upper limit
2. Subtract 0.5 from the lower limit

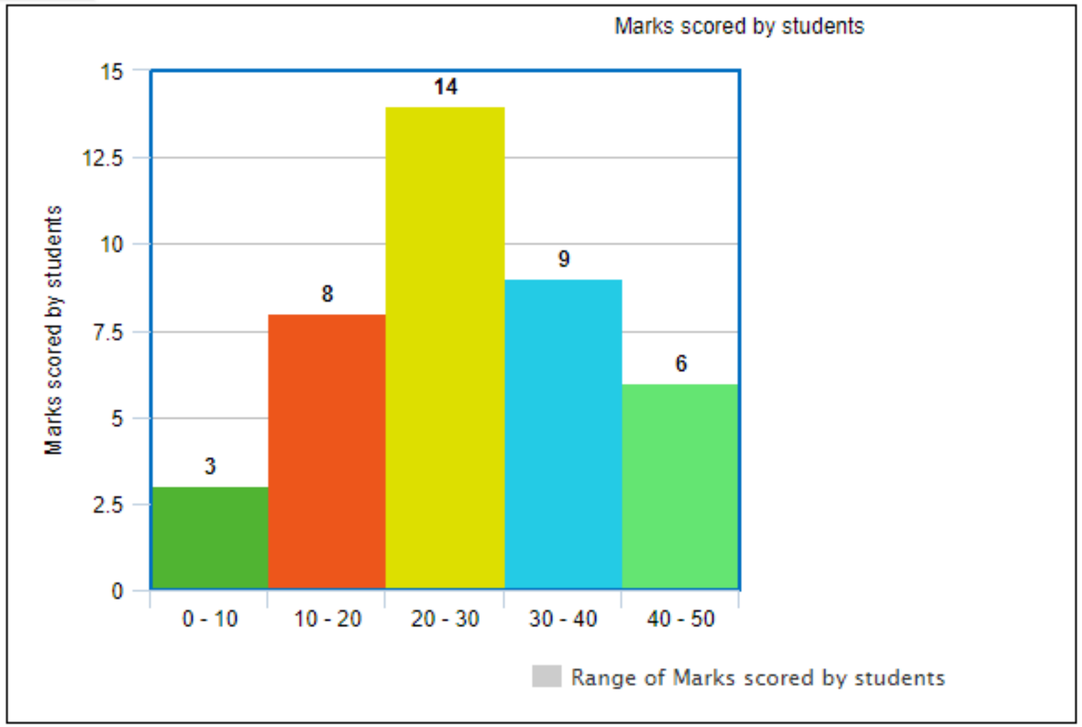
Example: The marks obtained by 40 students in an examination are given below

27	18	15	21	48	25	49	29	27	21	19	45	14	34	37	34	23	45	24	42
8	47	22	31	17	13	38	26	3	34	29	11	22	7	15	24	38	31	21	35

Prepare a histogram chart for the above data.

Weight (In Kg)	Number of Students
0-10	03
10-20	08
20-30	14
30-40	09
40-50	06
Total Students	60

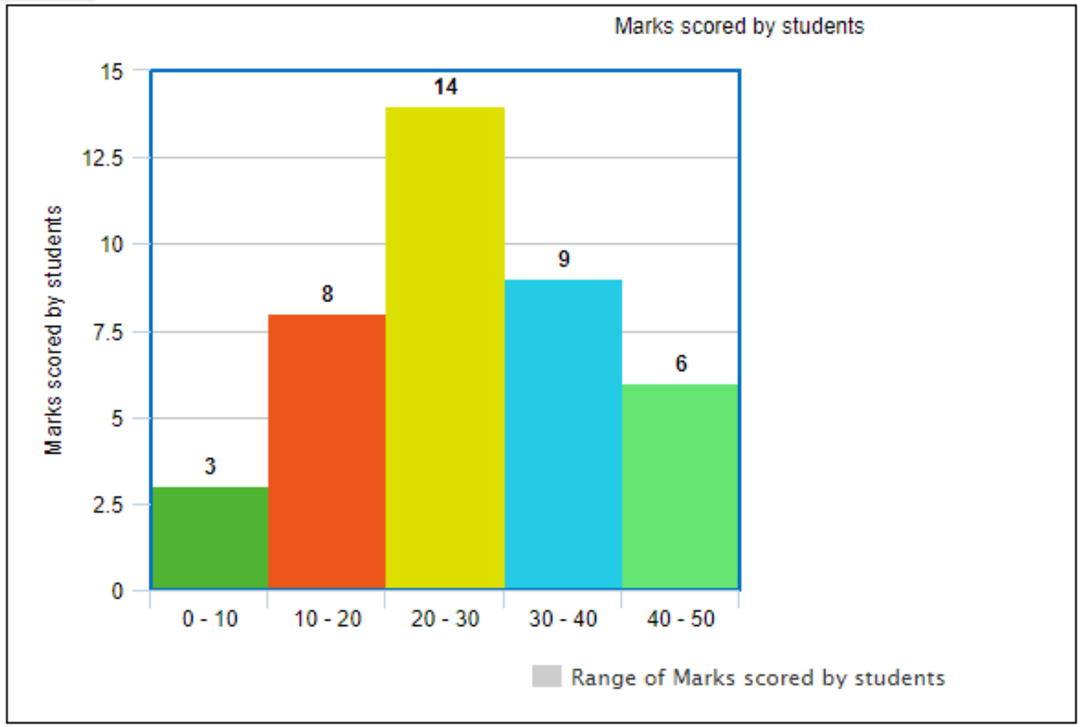
Univariate Visualizations



Interpretation:

- The histogram displays the frequency distribution of the weights of 40 students.
- The x-axis represents the ranges of weight categories.
- The y-axis represents the number of students falling within each weight category.

Univariate Visualizations



Interpretation:

- It shows that the majority of students fall within the weight range of 20-30 kg.
- There are fewer students with weights in the ranges of 0-10 kg and 40-50 kg.
- This histogram provides a clear visual representation of the weight distribution among the students in different weight categories.

Univariate Visualizations

Box plots: A boxplot presents data distribution using five summary statistics: minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum. It is commonly used to visualize the distribution, spread, and skewness of numerical data.

Here are the components of a boxplot:

Scale: Create an appropriate scale along the horizontal axis.

Box: Draw a box from the first quartile (Q1) to the third quartile (Q3), with a vertical line representing the median (Me).

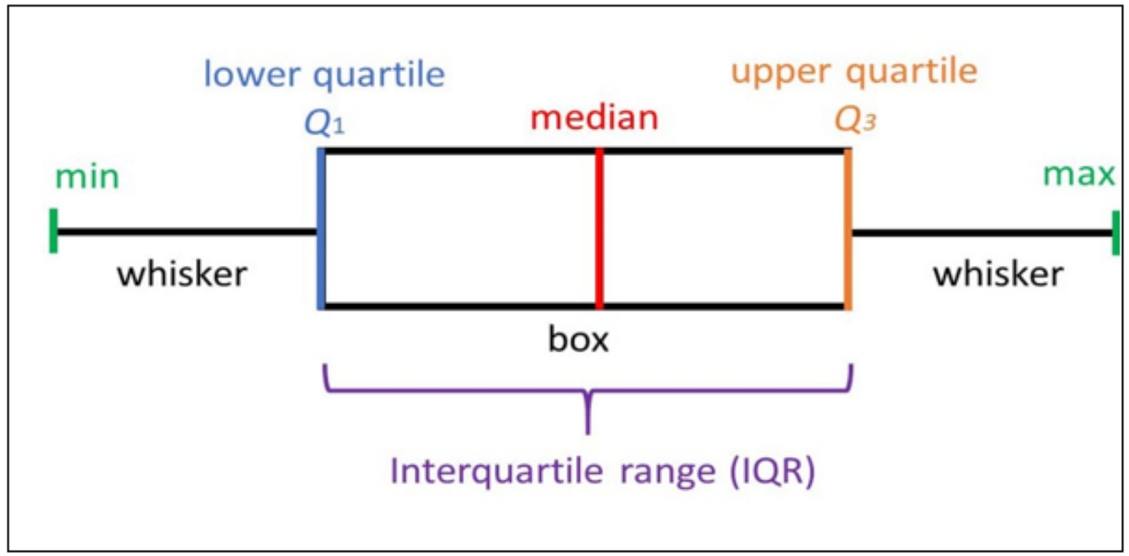
Inner and Outer Fences:

Determine the inner and outer fences: Determine the values of the inner fences and outer fences.

Univariate Visualizations

- The inner fence (Minimum), $Q_1 - 1.5 \times (IQR)$
- The upper inner fence (Maximum), $Q_3 + 1.5 \times (IQR)$
- The lower outer fence, $Q_1 - 3 \times (IQR)$ and
- The upper outer fence, $Q_3 + 3 \times (IQR)$

Whiskers: Draw dashed lines (whiskers) extending from Q1 to the first inner fence and from Q3 to the second inner fence.



Univariate Visualizations

Uses of Box Plot

A box plot provides us with the following information:

- By examining the relative position of the median line, we can guess the symmetry of the middle of the value.
- A measure of the variability of the values is given by the interquartile range.
- A box plot reveals that the distribution is positively or negatively skewed.
- Box plot is an easier alternative to forming a frequency distribution and plotting a histogram.

Univariate Visualizations

Example: Display the following data in a box plot,

12 5 6 12 6 14 16 12

Solution: By arranging the given data in ascending order, we have,

5 6 6 12 12 12 14 16

Here, the total number of observations, $n = 8\ (even)$

$$\text{Third quartile, } Q_3 = \frac{\frac{3n}{4}^{th}\ observation + (\frac{3n}{4}+1)^{th}\ observation}{2}$$

$$= \frac{\frac{3*8}{4}^{th}\ observation + (\frac{3*8}{4}+1)^{th}\ observation}{2}$$

$$= \frac{6^{th} observation + 7^{th}\ observation}{2}$$

$$= \frac{12+14}{2} = 13$$

Univariate Visualizations

$$\text{First quartile, } Q_1 = \frac{\frac{n}{4}^{th}\ observation + (\frac{n}{4}+1)^{th}\ observation}{2}$$

$$= \frac{\frac{8}{4}^{th}\ observation + (\frac{8}{4}+1)^{th}\ observation}{2}$$

$$= \frac{2^{th} observation + 3^{th}\ observation}{2}$$

$$= \frac{6+6}{2} = 6$$

Univariate Visualizations

- **Inter Quartile Range**, $IQR = Q_3 - Q_1 = 7$
- **Lower Inner fence**, $Q_1 - (1.5 * IQR) = -4.5$
- **Upper Inner fence**, $Q_3 + (1.5 * IQR) = 23.5$
- **Lower outer fence**, $Q_1 - (3 * IQR) = -15$
- **Upper outer fence**, $Q_3 + (3 * IQR) = 34$

Then draw a box-plot.

Bivariate Visualizations

Bivariate visualizations depict the relationship between two variables, often using scatter plots, heatmaps, or contour plots to show patterns or correlations between them

Scatter plots: Scatter plots are graphical representations of data points in a Cartesian coordinate system, where each point represents the value of two variables. They are useful for visualizing the relationship or correlation between the variables. It is used for construction and interpretation for visualizing the relationship between two continuous variables.

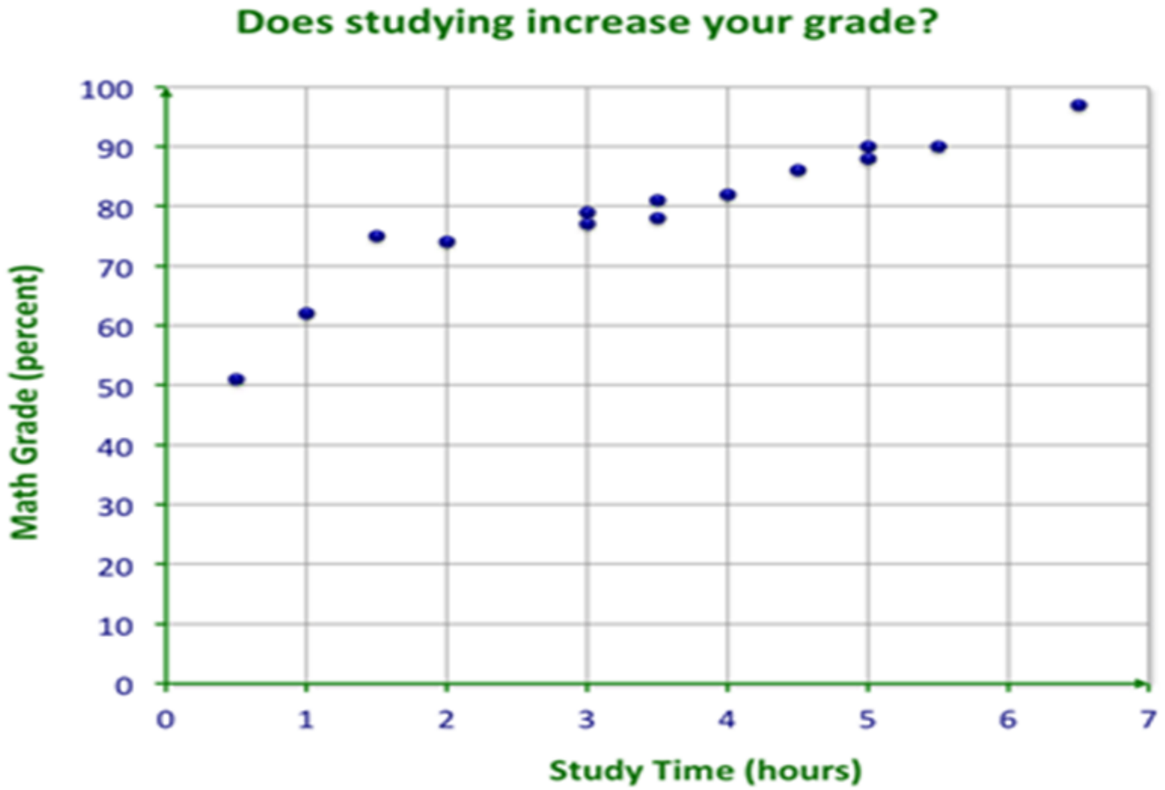
Example: Suppose, we have data on the daily study time of students and the percentage of marks obtained in the examination of 14 students.

Bivariate Visualizations

Student	Study Time (hour)	Marks in Percent
A	4.5	85%
B	4.0	82%
C	3.5	81%
D	3.5	78%
E	3	79%
F	3	78%
G	5	89%
H	5	90%
I	2	74%
J	6.5	98%
K	5.5	90%
L	1.5	70%
M	0.5	51%
N	1.0	62%

If we want to see the pattern of the change of grade and study time then we can use a scatterplot like following:

Bivariate Visualizations



Interpretation: You can see that there is a relationship between the daily study hour and math grades of the chart. The relationship is moving up to the right and therefore is a positive correlation.

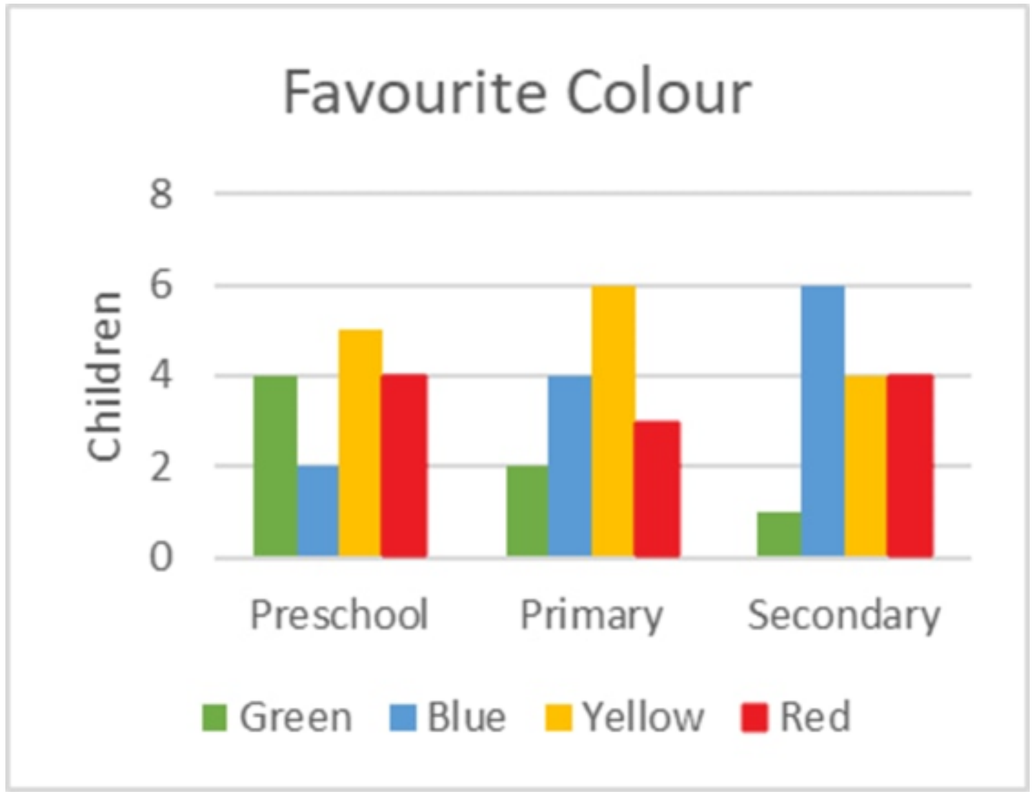
Bivariate Visualizations

Bar Plot:

A bivariate bar chart displays the relationship between two categorical variables by using bars to represent the frequency or proportion of each combination of categories from both variables. Each bar is segmented to show the distribution of one variable within each category of the other variable.

Example: Sometimes, when making a bar chart, we may realize that the information can not only be separated into categories, but into groups within them. In these situations, you could opt for a grouped bar chart. If we want to plot bar chart for the ‘favourite colour’ of 3 group (pre-primary, primary and secondary).

Bivariate Visualizations



Interpretation:

- 'Blue' is the most popular color for secondary students.
- Low popularity among preschool kids and average among primary children.
- Blue's popularity grows as children age.

Multivariate Visualizations

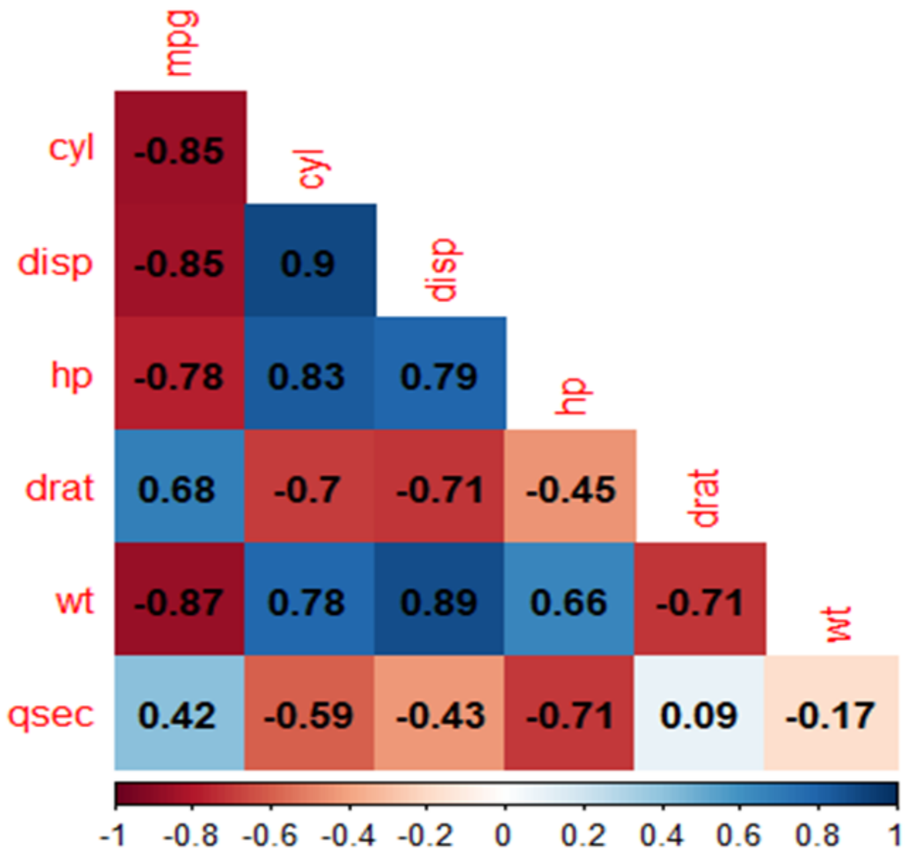
Heatmaps: Heatmaps visually represent data in a matrix format using color gradients, aiding in identifying patterns or correlations between variables. They are effective for displaying large datasets and are commonly used in data analysis and visualization. Used for construction and interpretation for displaying relationships between multiple variables using color intensity. This is used for:

- Quickly explore large datasets, identifying patterns and trends.
- Visualize correlations between variables for understanding relationships and dependencies.
- Facilitate decision-making in business strategy, marketing, and resource allocation.
- Monitor system performance in cybersecurity and network management.
- Aid in optimizing processes and systems by identifying areas of inefficiency or high performance.

Multivariate Visualizations

Interpretation:

- Dark blue shades signify strong positive correlations, where variables tend to change in tandem consistently.
- Dark red shades indicate the lowest correlations, suggesting weak or even negative relationships between variables.
- The gradient between these extremes showcases varying levels of correlation strength, with lighter shades representing moderate to weak correlations.



Multivariate Visualizations

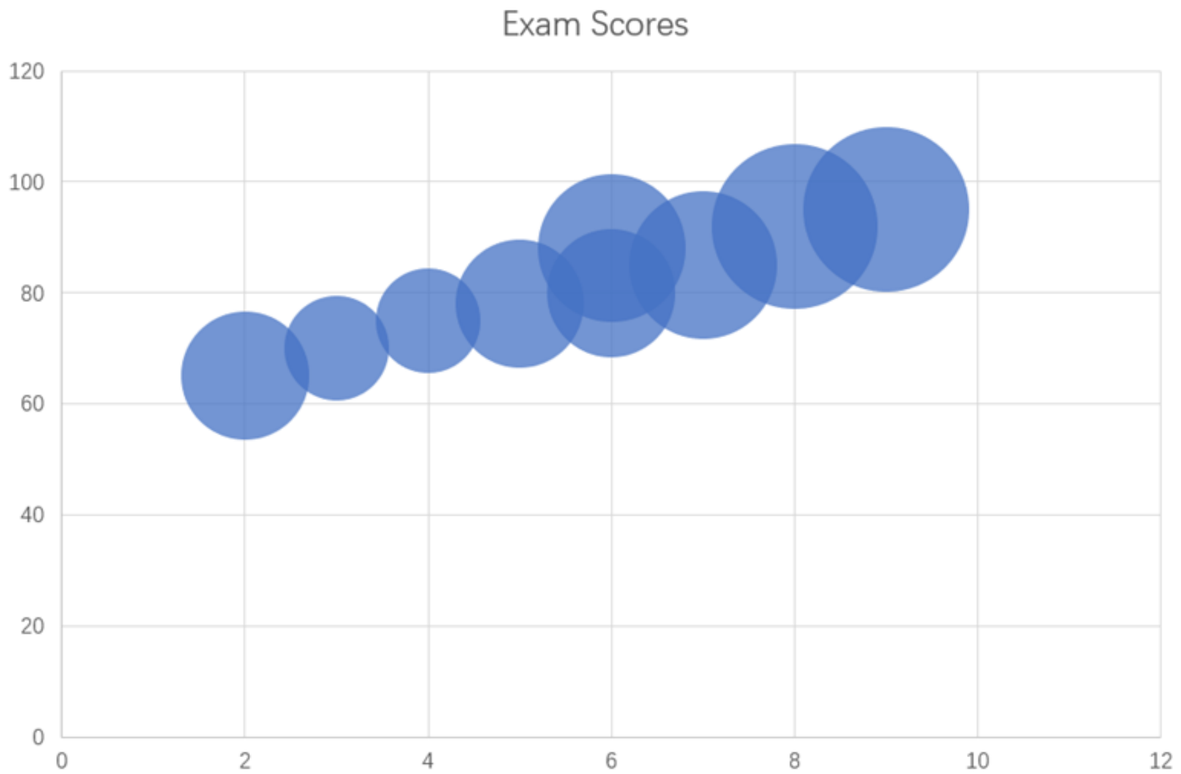
Bubble charts: Bubble charts visually represent data points on a graph with two axes, while the size of each bubble corresponds to a third numerical variable, providing insight into three dimensions of data simultaneously. They are useful for comparing relationships between multiple variables across different data points.

Example: Suppose a data given as,

Hours of Study	Exam Score	Difficulty Level
2	65	3
4	75	2
6	88	4
8	92	5
5	78	3
7	65	4
8	70	2
9	95	5
6	80	3

Multivariate Visualizations

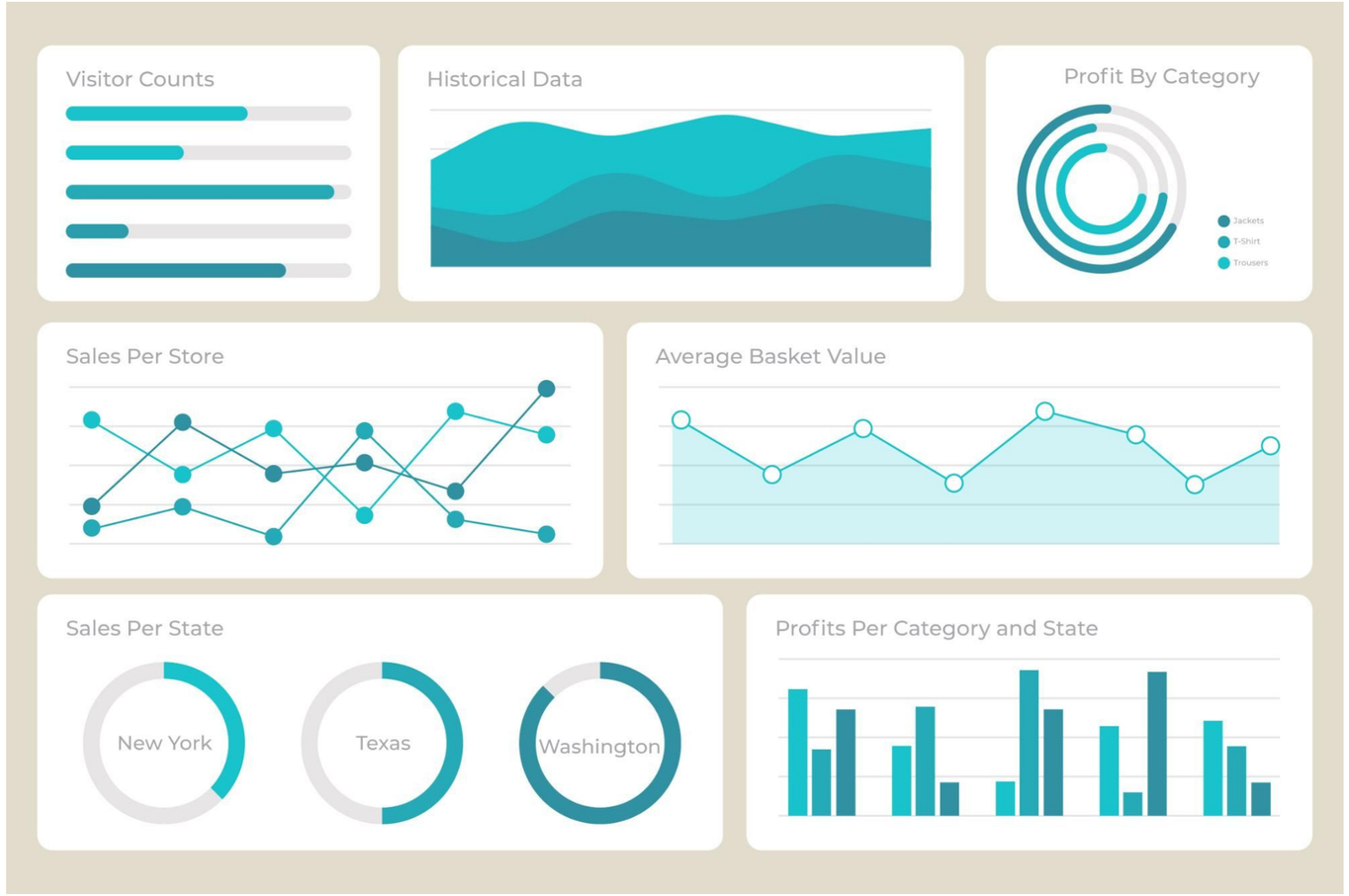
By constructing a bubble plot for exam score we find a representation like the following,



Interactive Visualizations

Interactive Visualizations: It enable users to manipulate and explore data dynamically through features like zooming, filtering, and tooltip interactions, enhancing data exploration and understanding.

- They facilitate immersive data analysis experiences, allowing users to uncover insights and patterns efficiently. (e.g., dashboards, interactive plots).
- Benefits of interactive visualizations for exploring and analyzing complex datasets.



Others types of Data Visualization

Line Plot: A line chart, or line graph, is a graphical representation used in elementary education to visually illustrate progression or pattern over time.

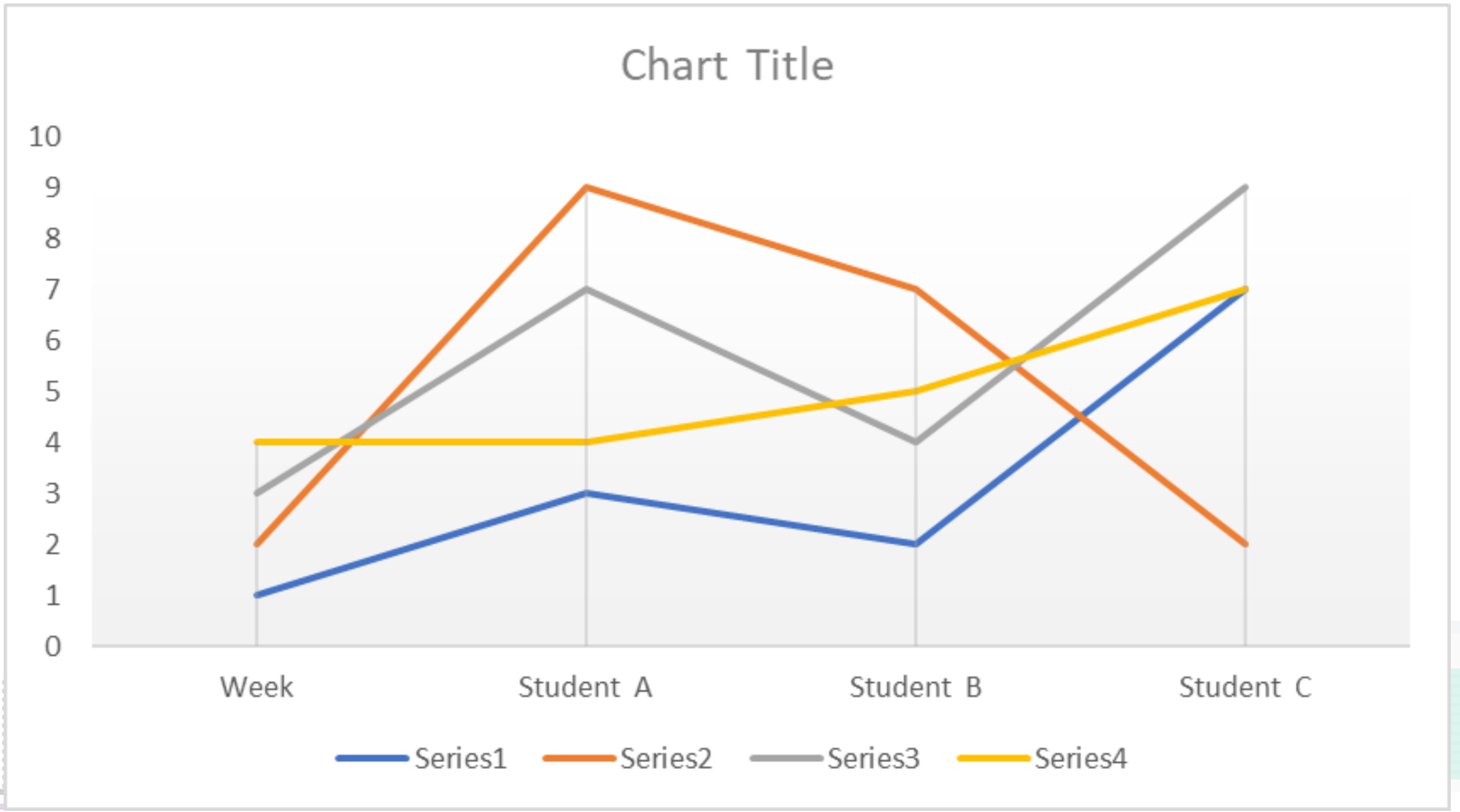
- Provide clear, visual data representation.
- Easy for young students to understand trends and changes.

Example: Let's say a primary school teacher wants to track the number of books read by each student in her class over the course of a month. She records the number of books read by each student each week and wants to visually represent this data to see if there are any trends in reading habits. She can create a line chart with the following data:

Week	Student A	Student B	Student C
1	3	2	4
2	4	3	5
3	5	4	6
4	6	5	7

Others types of Data Visualization

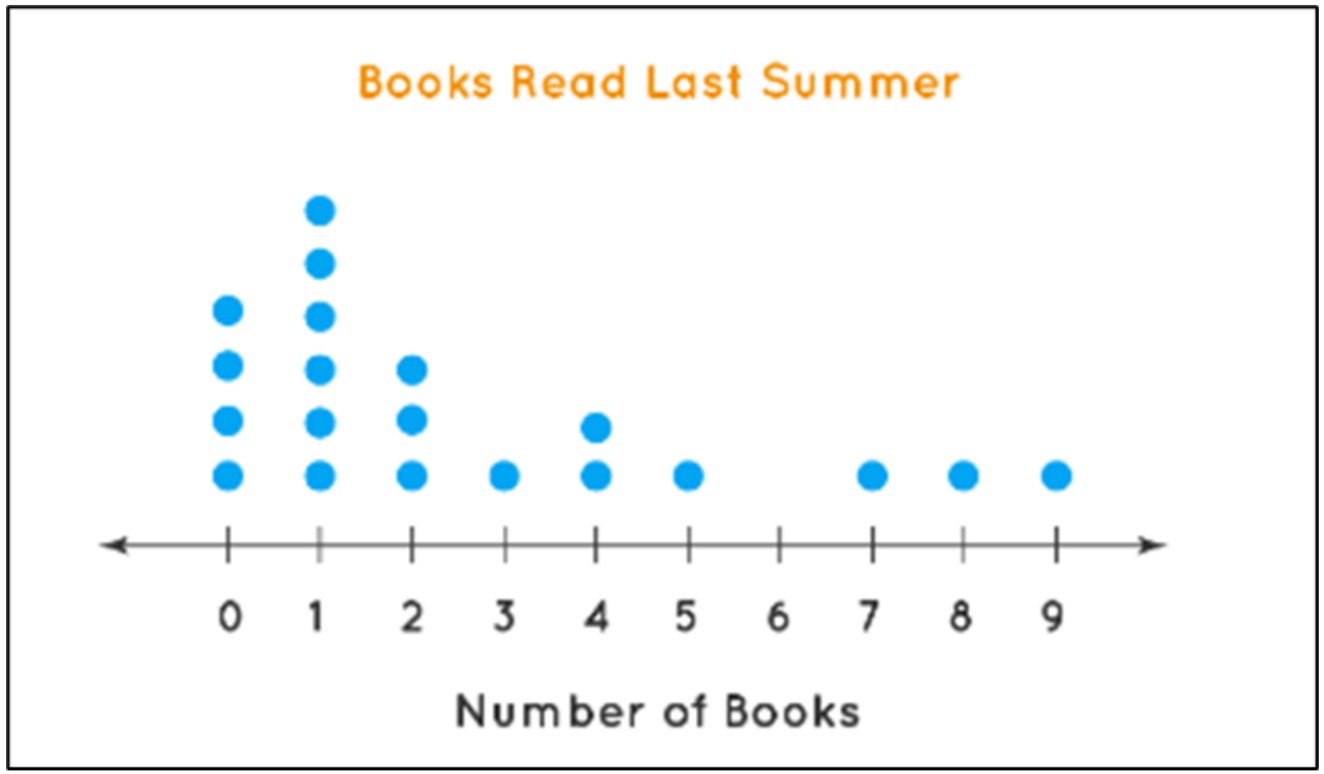
In this example, each student's reading progress is represented by a line on the chart, with the x-axis showing the weeks and the y-axis showing the number of books read. The teacher can quickly see how each student's reading habits change over the four-week period and identify any patterns or differences between students.



Others types of Data Visualization

Dot Plot: A dot plot is a graph that displays data using dots placed along an axis. Each dot represents a single data point, and the position of the dot along the axis corresponds to the value of that data point.

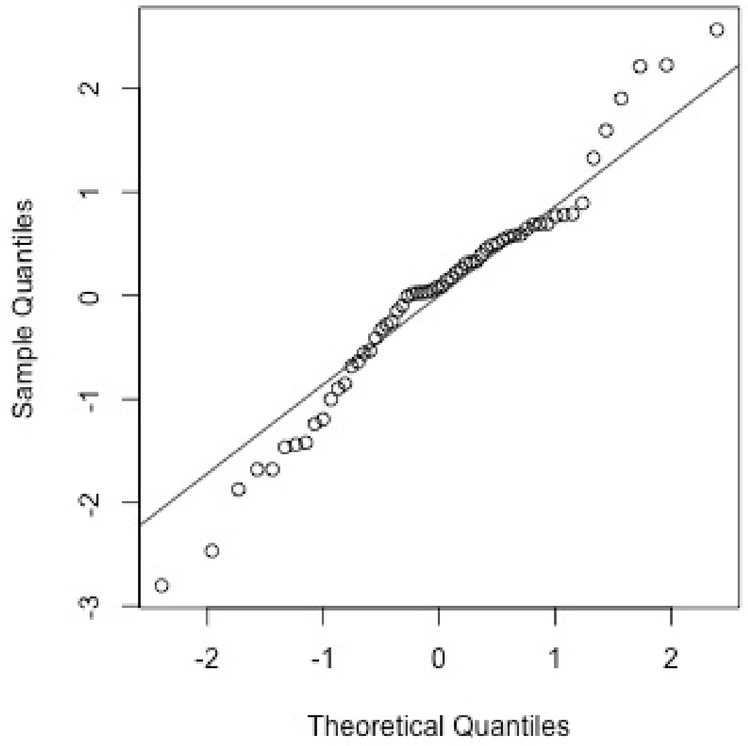
Example: Let's say a primary school teacher wants to visualize the number of books read in last summer by the students in her class. She can create a dot plot using the following data:
0 0 0 0 1 1 1 1 1 1 2 2 2 3 4 4 5 7 8 9



Others types of Data Visualization

Q-Q Plot: A Q-Q plot assesses if a set of data follows a normal distribution by comparing the quantiles of the dataset to the quantiles of the theoretical distribution. If the dataset follows the theoretical distribution, the plot points align.

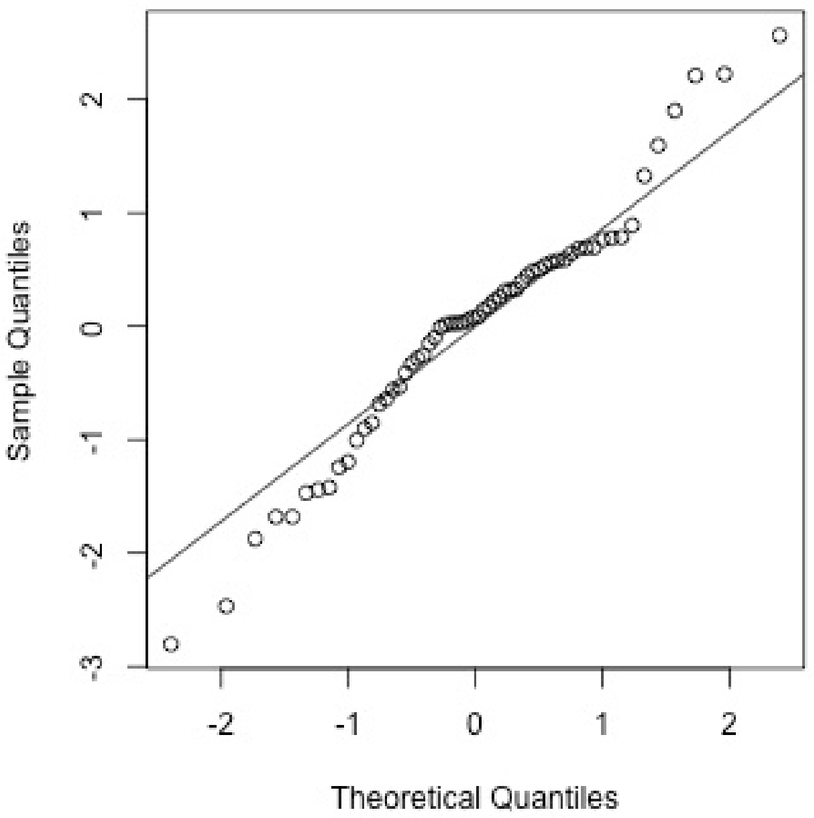
Example: Let's consider a simplified example involving the heights of students in a primary school class. Suppose a teacher wants to assess if the heights of students in the class follow a normal distribution.



Others types of Data Visualization

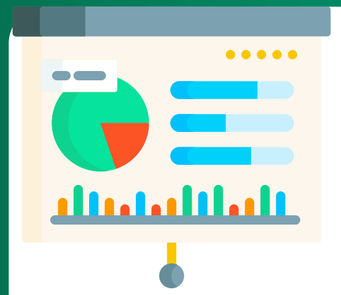
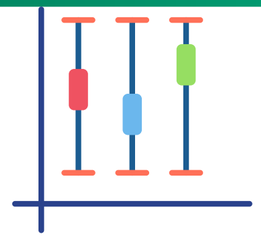
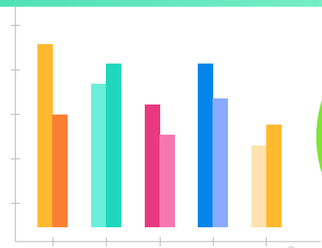
Interpreting the Q-Q Plot:

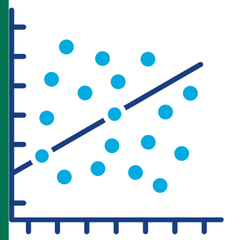
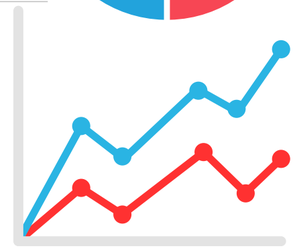
- If the data closely follows a normal distribution, the points on the Q-Q plot will fall approximately along a straight line.
- Deviations from a straight line indicate departures from normality.



Misleading Visualizations

- Discussion on ethical considerations related to creating visualizations that may be misleading or misinterpreted

Thank You