

# Multi Disciplinary Course (MDC) *Mathematical Science*

## **MDC-1: Foundation Of Mathematical Science -- I**

*For FYUGP 1<sup>st</sup> Semester (Unit III -- Statistics)*

By

**R K Bhuyan,**

HoD, DoM & Vic-Principal

Tihu College, Tihu

Class No 1

# *Historical Development of Statistics*

The historical development of statistics is a fascinating journey that spans thousands of years, evolving from simple record-keeping to a sophisticated field essential to science, business, policy, and everyday decision-making.

Here's a concise overview divided into major periods:



# Summary Timeline

## Period

Ancient

Renaissance(1500–1700)

18th century

19th century

Early 20th century

Mid 20th century

Late 20th century

21st century

## Key Features

Censuses, record-keeping

Political arithmetic, demographic studies

Birth of probability theory, statistical thinking

Statistical methods for society, correlation, regression

Modern inferential statistics, experimental design

Hypothesis testing, statistical theory

Computer-based statistics, Bayesian resurgence

Big data, data science, AI, machine learning

# 1. Ancient Times: Origins of Statistics (Before 500 AD)

- **Early Uses:** The word “statistics” comes from the **Latin *status*** (state). In ancient civilizations like **Babylonia, Egypt, China, and Rome, Bharat** -rulers collected data on **population, land, and taxes**—primitive forms of what we now call descriptive statistics.
- **Censuses:** One of the oldest known statistical tools. For example:
  - **Egyptians collected census data for building pyramids (~3000 BCE).**
  - **Romans used censuses for military and tax purposes.**

## 5. 19th Century: Emergence as a Discipline

- **Adolphe Quetelet** introduced the idea of the “average man” and applied statistical methods to social sciences.
- **Francis Galton** and **Karl Pearson** developed methods like *correlation* and *regression*. Pearson also founded the first university statistics department.
- **Charles Babbage** promoted data collection and statistical thinking for industrial and scientific purposes.
- Statistics began to be recognized as an **independent field**, especially in England and Germany.

## 6. Early 20th Century: Formalization and Expansion

- **Ronald A. Fisher** revolutionized experimental design and inference with works like *"Statistical Methods for Research Workers"* (1925).
- Developed concepts of:
  - **Analysis of Variance (ANOVA)**
  - **Maximum Likelihood Estimation**
- **Jerzy Neyman** and **Egon Pearson** developed the framework for **hypothesis testing** (Neyman-Pearson Lemma).
- Growth of statistical applications in **agriculture, biology, psychology, and education.**

## 7. **Mid to Late 20th Century: Computational Revolution**

- Widespread use of **computers** transformed statistical practice:
  - Enabled more complex models (e.g., logistic regression, time series).
  - Growth of **Bayesian methods**, simulation (e.g., **Monte Carlo**), and machine learning.
- Development of statistical software: **SPSS, SAS, R, Stata**, etc.
- Expansion into new fields: **Economics, Epidemiology, Social Sciences, and Environmental science.**

## 8. **21st Century: Data Science Era**

- Explosion of “**big data**” due to the internet, sensors, and digital records.
- **Statistics merged with computer science to form data science.**
- Increased focus on:
  - **Predictive modelling**
  - **Machine learning & AI**
  - **Bayesian methods**
  - **Causal inference**
- Ethical concerns and reproducibility issues brought renewed interest in transparency and robustness.

### **Q1: What is the origin of the word “statistics”?**

- **A:** The word “statistics” comes from the Latin word *status*, meaning “state.” It originally referred to data collected for state governance, such as censuses and tax records.

### **Q2: Which ancient civilizations used early forms of statistics, and for what purpose?**

- **A:** Ancient civilizations like **Egypt, Babylon, China, and Rome** used statistics primarily for **population censuses, taxation, and resource planning.**

### **Q3: Who is considered the father of demography, and what was his contribution?**

- **A: John Graunt** is considered the father of demography. In 1662, he analysed London’s Bills of Mortality, identifying patterns in birth and death rates—an early use of statistical analysis in public health.

## Q4: How did the development of probability theory influence statistics?

- **A:** Probability theory provided the **mathematical foundation** for statistical inference. It began with gambling problems and evolved with contributions from **Jacob Bernoulli**, **Bayes**, and **Laplace**, enabling predictions based on incomplete data.

## Q5: What were Karl Pearson's major contributions to statistics?

- **A:** Karl Pearson developed key statistical tools such as the **correlation coefficient**, **chi-square test**, and **regression analysis**. He also founded the world's first statistics department at University College London.

## Q6: How did computers influence the development of statistics in the late 20th century?

- **A:** Computers enabled the analysis of large and complex datasets, making possible the development of **simulation techniques, machine learning, and advanced modelling**, and also led to the creation of **statistical software** like **SPSS, SAS, and R**.

## Q7: What is the relationship between statistics and data science in the 21st century?

- **A:** Statistics is a core component of **data science**, which combines statistical methods with **computer science** and **domain knowledge** to extract insights from large-scale and complex data.

## Q8: What modern challenges and focuses exist in the field of statistics today?

- **A:** Modern statistics deals with **big data, ethical concerns, reproducibility, AI**, and the need for **causal inference**, all while trying to maintain rigor and transparency in data analysis.

# Why Is Statistics Important in Everyday Life?

Purpose	How Statistics Help
Decision Making	Helps us choose the best option (e.g., cheapest phone plan, safest route)
Risk Assessment	Helps assess the chances of an event happening (e.g., catching a disease, winning a lottery)
Understanding Trends	Allows us to track changes over time (e.g., inflation, climate change)
Making Predictions	Based on past data, we can predict future outcomes (e.g., sales trends, exam scores)

# Statistics in Everyday Life

## 1. Health and Medicine

- **Doctors** use statistics to recommend treatments (e.g. “This drug has a 90% success rate”).
- **COVID-19 data**, like infection rates and vaccine effectiveness, are based on statistical analysis.
- Health apps track our steps, calories, and heart rate—then use stats to provide insights.

## 2. Weather Forecasting

- Forecasts like “60% chance of rain tomorrow” come from statistical models analysing historical and current weather data.

## 3. Finance and Budgeting

- **Banks** assess credit scores using statistical models.
- **Investors** use stats to analyse stock trends and risks.
- **Personal budgeting** apps analyse our spending habits and predict future expenses.

# Statistics in Everyday Life

## 4. Education

- Schools use statistics to evaluate student performance (e.g. exam averages, pass rates).
- Governments use statistical data to improve education policy and funding allocation.

## 5. Shopping and Marketing

- **Retailers** track customer behaviour using statistics (e.g., what time people shop, which products are popular).
- Personalized ads are shown based on statistical analysis of our browsing and buying history.

## 6. Social Media

- Platforms like Instagram and YouTube use algorithms based on statistics to suggest content.
- Metrics such as likes, shares, and engagement rates are analysed statistically.

# Statistics in Everyday Life

## 7. Transportation

- **Google Maps** and ride apps use traffic data and statistical models to estimate travel time.
- Airlines use statistics for overbooking decisions and flight scheduling.

## 8. Sports

- Teams use **statistics to track player performance**, plan strategies, and predict opponents' moves.
- Fans see stats like win probabilities and player rankings.

# Questions and Answers

## ◆ Que1: What is statistics in simple terms?

- **Ans:** Statistics is the science of collecting, analysing, and interpreting data to make decisions or understand patterns.

## ◆ Que2: How are statistics used in weather forecasting?

- **Ans:** Meteorologists use statistical models to analyse past and current weather data to predict future conditions, like saying there's a "60% chance of rain."

## ◆ Que3: How do doctors use statistics in medicine?

- **Ans:** Doctors use statistics to evaluate treatment effectiveness, risks, and survival rates. For example, a drug might be said to have a 95% success rate based on clinical trial data.

◆ **Que4: What role do statistics play in budgeting and personal finance?**

- **Ans:** Budgeting apps use statistics to analyze spending habits, predict future expenses, and help users make informed financial decisions.

◆ **Que5: Give an example of statistics used in education.**

- **Ans:** Schools use student test scores and grade averages to assess performance and identify areas for improvement. Governments use education statistics for policy decisions.

◆ **Que6: How are statistics used in social media platforms?**

- **Ans:** Platforms like YouTube and Instagram use statistical algorithms to recommend content, track user engagement, and measure popularity through likes, views, and shares.

### ◆ **Que7: In what way do retailers use statistics?**

- **Ans:** Retailers use customer purchase data to track trends, personalize ads, and manage inventory based on consumer behavior patterns.

### ◆ **Que8: What's a real-life example of statistics in transportation?**

- **Ans:** Google Maps uses real-time traffic data and past travel patterns to estimate travel times and suggest the fastest routes.

### ◆ **Que9: How are statistics used in sports?**

- **Ans:** Coaches and analysts use player statistics (e.g., goals scored, passing accuracy) to plan strategies and evaluate performance.

### ◆ **Que10: Why is understanding statistics important in everyday life?**

- **Ans:** Understanding statistics helps us make smarter decisions, evaluate risks, interpret news, and understand trends in health, finance, education, and more.

# Multi Disciplinary Course (MDC) *Mathematical Science*

**MDC-1: Foundation Of Mathematical Science -- I**  
*For FYUGP 1<sup>st</sup> Semester (Unit III -- Statistics)*

By

**R K Bhuyan,**

HoD, DoM & Vic-Principal

Tihu College, Tihu

# Scope and Importance of Statistics

Statistics is a branch of mathematics that deals with the collection, analysis, interpretation, and presentation of data. It provides the tools and methods necessary to make sense of large amounts of information and draw meaningful conclusions. Its scope is vast, and its applications are found in nearly every field of study.

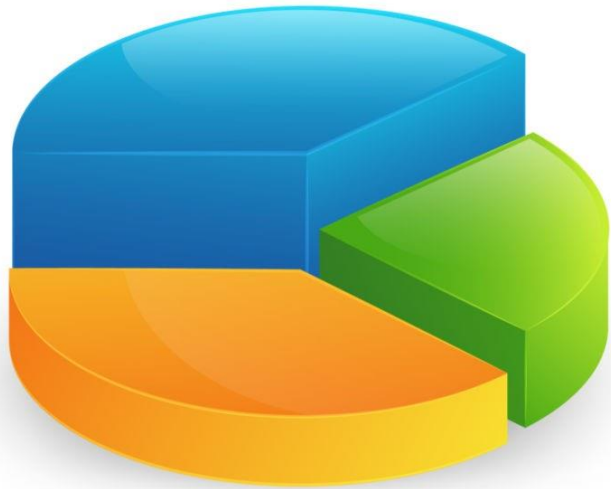
# Scope and Importance

The scope of statistics is broad, and it is a crucial tool in modern decision-making, research, and policy formulation.

It allows us to:

- **Organize and summarize data:** Statistics helps to condense large datasets into manageable and understandable summaries, like averages, percentages, and graphs.

# 3D GRAPH



PIE CHART



BAR CHART

- **Draw inferences:** It enables us to make educated guesses or predictions about a larger population based on a smaller, representative sample. This is fundamental to scientific research and market analysis.
- **Identify relationships and trends:** By analyzing data, statisticians can uncover hidden patterns, correlations, and trends that can inform business strategy, public health initiatives, and economic forecasts.
- **Test hypotheses:** It provides a structured, scientific method for testing theories and claims, such as the effectiveness of a new drug or the impact of a new government policy.

# Limitations 🧐

Despite its immense utility, statistics has several key limitations that must be understood to avoid misuse and misinterpretation:

- **It only deals with quantitative data:** Statistics cannot directly analyze qualitative characteristics like honesty, beauty, or intelligence unless they are converted into a numerical format, such as through a scoring system.
- **It doesn't study individual cases:** Statistical conclusions are based on averages and aggregates of facts, not on specific individual instances. The average height of people in a country doesn't tell you the height of any one person.

# Limitations 🧐

- **Results are true only on average:** Statistical laws are not universally true like those in physics or chemistry. They are based on probability and are true in the long run or on average, which means there's always a chance for variation.
- **It can be misused:** Statistics can be manipulated or presented in a misleading way to support a desired conclusion, often without providing the full context. This can lead to false or fallacious arguments

# Applications in Other Fields

Statistics is an indispensable tool in a wide range of fields, including:

- **Business and Economics:** Used for market research, financial forecasting, quality control, and risk management. Businesses use statistics to understand consumer behavior and make data-driven decisions about products and marketing.
- **Healthcare and Medicine:** Essential for clinical trials to test the efficacy of new drugs and treatments. It's also used in epidemiology to study the spread of diseases, assess public health risks, and manage patient data.
- **Government and Public Policy:** Governments rely on statistics from sources like censuses to make informed decisions about resource allocation, infrastructure, and social programs. It's used to analyze population growth, unemployment rates, and inflation.

# Applications in Other Fields

- **Science and Research:** Statistics is the backbone of scientific research. It helps scientists design experiments, analyze experimental data, and draw valid conclusions. This applies to fields from biology and chemistry to astronomy and environmental science.
- **Social Sciences:** Used in sociology and psychology to analyze human behavior, social trends, and survey data. It helps researchers understand complex social phenomena, such as public opinion on a political issue or the factors contributing to educational success.
- **Sports:** Professional sports teams use statistical analysis (like Sabermetrics in baseball) to evaluate player performance, develop game strategies, and inform recruitment decisions.

# Roles of computers in statistic

Computers play an indispensable role in modern statistics by automating and accelerating the entire data analysis process. They have transformed the field from manual, labour-intensive calculations to powerful, data-driven insights.

# Primary Roles of Computers in Statistics

- **Data Handling and Management:** Computers are essential for **collecting, organizing, and storing** vast amounts of data. Databases and spreadsheets allow statisticians to manage large, complex datasets that would be impossible to handle manually.
- **Performing Complex Calculations:** Computers can perform complex statistical calculations with incredible speed and accuracy. **Regression analysis, hypothesis testing, and time series analysis** are more advanced statistics.

- **Data Visualization:** Computers enable the creation of sophisticated and dynamic visualizations. Statistical software can generate a wide array of charts and graphs, including scatter plots, histograms, and heat maps. These visual representations help statisticians and others to easily identify patterns, trends, and outliers in the data, making it easier to communicate findings.
- **Simulation and Modeling:** Computers are crucial for running **simulations** and creating **statistical models**. These simulations are used to model complex systems, from financial markets to disease spread.
- **Automation of Repetitive Tasks:** With programming languages like R and Python, statisticians can **automate repetitive analytical tasks**. This ensures reproducibility and reduces the chance of human error.

# Statistical Software and Programming Languages

A wide range of software and programming languages has been developed to leverage the power of computers for statistical work:

- **R and Python:** These are open-source programming languages with extensive libraries for statistical analysis (e.g., NumPy, Pandas, Scikit-learn in Python; ggplot2, tidyverse in R). They are highly popular in academia and data science for their flexibility and power.
- **SPSS (Statistical Package for the Social Sciences):** A commercial software widely used in the social sciences for its user-friendly, graphical interface, which simplifies data analysis for those who aren't experts in coding.
- **SAS (Statistical Analysis System):** Another powerful commercial software used heavily in business and government for its robust data management and advanced analytical capabilities.
- **Stata:** A commercial statistical software package primarily used in economics, political science, and biomedical research, known for its extensive set of built-in commands and easy-to-use interface.

# Multi Disciplinary Course (MDC) *Mathematical Science*

**MDC-1: Foundation Of Mathematical Science -- I**  
*For FYUGP 1<sup>st</sup> Semester (Unit III -- Statistics)*

By

**R K Bhuyan,**

HoD, DoM & Vic-Principal

Tihu College, Tihu

# Statistical Data Types

Statistical data can be broadly classified into two main types:

**Qualitative** :- Qualitative data is non-numerical. *Examples include gender (male, female), marital status (single, married), or blood type (A, B, AB, O).*

**and**

**Quantitative** :- while quantitative data is numerical. *Examples include the number of students in a class or the number of cars in a parking lot.*

# Methods of Data Collection

Data collection is the process of gathering and measuring information from various sources to get a complete picture of a specific topic. The two main types of data collection are primary and secondary.

- **Primary Data Collection:** This is the process of collecting original, firsthand data specifically for your research purpose.
- **Secondary Data Collection:** This involves using data that has already been collected and published by someone else for a different purpose.

# Primary Data Collection:

- **Surveys and Questionnaires:** One of the most common methods, used to collect data by asking a series of questions to a group of people. This can be done online, in person, or by mail.
- **Interviews:** Involve direct, one-on-one conversations with individuals to gather in-depth, qualitative information about their opinions, experiences, or knowledge.
- **Observations:** The researcher observes and records behaviours or events as they naturally occur in a specific setting. This method is useful for studying human behaviour or processes.
- **Experiments:** A controlled method where researchers manipulate one or more variables to see how they affect others. This is common in scientific research to test hypotheses and establish cause-and-effect relationships

# Secondary Data Collection:

- **Published Sources:** This includes data found in government publications (like census reports), academic journals, books, and public company records.
- **Internal Sources:** This refers to data that is already within an organization, such as sales records, customer transaction data, and employee reports.
- **External Sources:** This includes data from third-party sources, like market research reports, professional organizations, and trade publications.

<https://youtu.be/OeF93Q8gPsM>

# General Concepts

## Que1: What is statistics?

**Ans:** Statistics is the science of collecting, analyzing, interpreting, and presenting data. It helps us make sense of information and draw conclusions about a larger population based on a smaller sample.

## Que2: What is the main difference between qualitative and quantitative data?

**Ans:** **Qualitative data** is non-numerical and describes characteristics (like colors or opinions), while **quantitative data** is numerical and represents counts or measurements (like height or temperature).

# Scope and Limitations

**Que3: What are two key applications of statistics in the real world?**

**Ans:** Statistics is used in:

- 1.Business** to forecast sales and analyze market trends.
- 2.Medicine** to test the effectiveness of new drugs through clinical trials.

**Que4: What is a major limitation of statistics?**

**Ans:** A major limitation is that statistical conclusions are based on averages and may not apply to any individual case. It also cannot analyze subjective qualities unless they are converted into numbers.

# Data Types and Collection

**Que5: What is the difference between discrete and continuous data?**

**Ans: Discrete data** consists of whole numbers that are counted (e.g., number of students).

**Continuous data** can take any value within a range and is measured (e.g., a person's height).

**Que6: What is the difference between primary and secondary data collection?**

**Ans: Primary data** is original, first-hand data collected specifically for a research purpose (e.g., conducting a survey).

**Secondary data** is data that has already been collected by someone else and is being used again (e.g., using government census data).

# Role of Computers

**Que7: How have computers changed the field of statistics?**

**Ans:** Computers have revolutionized statistics by allowing for the **fast and accurate analysis** of massive datasets, automating complex calculations, and enabling sophisticated data visualization.

**Que8: Name two common software tools used for statistical analysis.**

**Ans: R** and **Python** are widely used for statistical analysis due to their powerful libraries and capabilities. Other tools include SPSS and SAS, which are often used in business and social sciences.

# Multi Disciplinary Course (MDC)

## *Mathematical Science*

### **MDC-1: Foundation Of Mathematical Science -- I**

*For FYUGP 1<sup>st</sup> Semester (Unit III -- Statistics)*

By

**R K Bhuyan,**

HoD, DoM & Vic-Principal

Tihu College, Tihu

Class No 4

# Time Series Data

## Definition:

Time series data is a sequence of data points collected or recorded at specific time intervals (e.g., hourly, daily, monthly, annually).

## Key Features:

- Time is a crucial component.
- Data is ordered chronologically.
- Used to analyse trends, patterns, or cycles over time.

## Examples:

- Daily stock prices.
- Monthly rainfall in a city.
- Annual GDP of a country.
- Temperature readings every hour.

# Qualitative Data *(also called Categorical Data)*

## Definition:

Qualitative data describes qualities or characteristics and cannot be measured in numbers. It represents categories or labels.

## Types:

- **Nominal:** Categories with no natural order (e.g., gender, eye color, nationality).
- **Ordinal:** Categories with a meaningful order, but not evenly spaced (e.g., rating levels like “poor,” “fair,” “good”).

## Examples:

- Hair color: black, brown, blonde.
- Survey response: agree, neutral, disagree.
- Blood type: A, B, AB, O.

# Quantitative Data

## Definition:

Quantitative data consists of numeric values that represent quantities and can be measured or counted.

## Types:

- **Discrete:** Countable values (e.g., number of students in a class).
- **Continuous:** Measurable quantities that can take any value within a range (e.g., height, weight, temperature).

## Examples:

- Age in years.
- Income in dollars.
- Number of books on a shelf.
- Time taken to complete a task.

# Quick Comparison Table

Type	What It Describes	Examples
Time Series Data	Data over time	Stock prices over months
Qualitative Data	Categories or qualities	Eye colour, gender, car brand
Quantitative Data	Numeric measurements or counts	Age, weight, income, test scores

### **Que1: What is time series data?**

**Ans:** Time series data is a set of observations recorded at specific time intervals, typically ordered chronologically. It is used to track changes or trends over time.

### **Que2: Give an example of time series data.**

**Ans:** An example of time series data is the daily closing price of a company's stock over one year.

### **Que3: What is the main purpose of analysing time series data?**

**Ans:** The main purpose is to identify trends, seasonal patterns, and cyclical behaviour, and to make forecasts or predictions based on past data.

#### **Que4: What are the components of time series data?**

**Ans:** The main components are:

- **Trend** (long-term increase/decrease),
- **Seasonality** (regular pattern over short time periods),
- **Cyclicity** (long-term up-and-down movements),
- **Irregular or random variations** (unexpected fluctuations)

#### **Que5 : What are common methods used in time series analysis?**

**Ans:** Some common methods include:

- Moving averages
- Exponential smoothing
- ARIMA models (Auto Regressive Integrated Moving Average)
- Decomposition methods
- Seasonal adjustment

#### **Que6: What tools or software are commonly used for time series analysis?**

**Ans :** Common tools include:

- **Excel**
- **R (with packages like forecast)**
- **Python (with libraries like pandas, stats models, prophet)**
- **SPSS, SAS, MATLAB**

# Multi Disciplinary Course (MDC) *Mathematical Science*

**MDC-1: Foundation Of Mathematical Science -- I**  
*For FYUGP 1<sup>st</sup> Semester (Unit III -- Statistics)*

By

**R K Bhuyan,**  
HoD, DoM & Vic-Principal  
Tihu College, Tihu

Class No 5

# Data Representation in Statistics

**Data representation** in statistics refers to the different **ways of organizing, summarizing, and displaying data** so it can be easily understood and analysed. It helps in identifying patterns, trends, and outliers in data.

# Types of Data Representation

## Tabular Representation

- Data is organized in rows and columns.
- Makes it easy to compare and look up values.

### Example:

Student

Ram

Hari

Jadu

Madhu

Marks

85

78

90

56

# Textual Representation

Data is described in words or sentences.

- Used when data is small and simple.

**Example:**

"Out of 100 students, 60 passed and 40 failed."

# Diagrammatic Representation

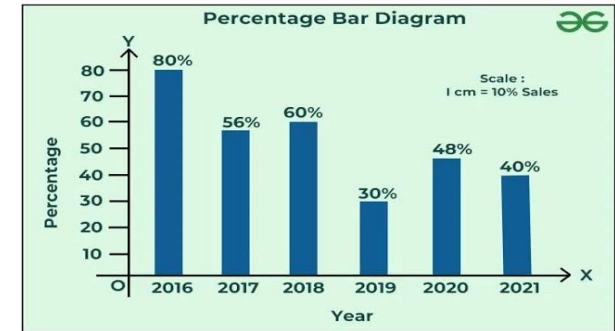
Visual representation using diagrams or charts. Common types:

## 1. Bar Graph / Bar Chart

Used for **categorical or discrete data**.

Each bar represents a category.

*Example: Population by country.*

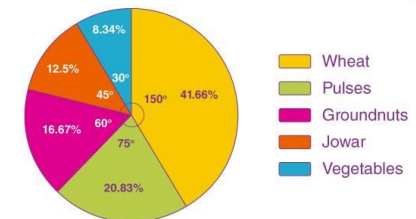


## 2. Pie Chart

Represents **percentages or proportions**.

The whole circle = 100%.

*Example: Market share of different companies.*

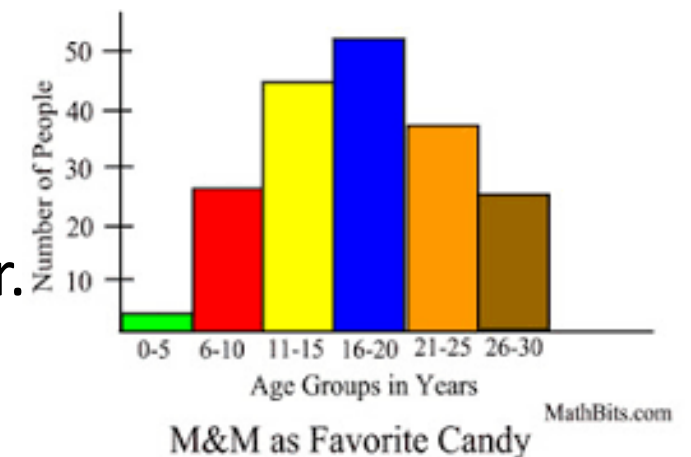


## 3. Histogram

Used for **continuous quantitative data**.

Similar to bar charts but bars touch each other.

*Example: Distribution of ages in a group.*

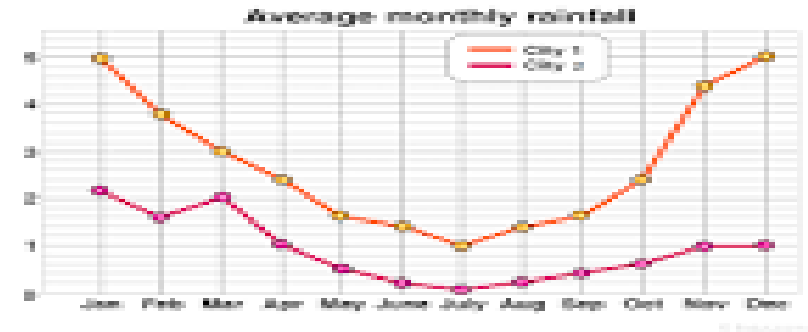


## 4. Line Graph

Used for **time series data**.

Shows trends over time.

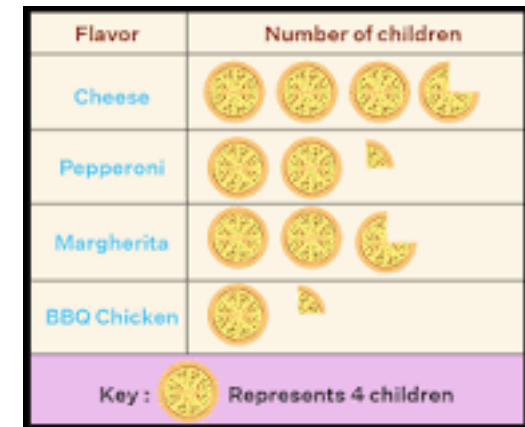
*Example: Temperature over a week.*



## 5. Pictogram / Pictograph

Uses pictures or symbols to represent data.

Easy to understand for children or general audiences.



# Graphical Representation in Statistics

Graphical tools provide a more detailed and visual insight into data, especially large datasets. Common statistical graphs include:

Graph	Used For
Line Graph	Time series data
Histogram	Frequency of continuous data
Box Plot (Box-and-whisker)	Spread and outliers in data
Scatter Plot	Relationship between two variables
Frequency Polygon	Comparison of distributions

# Why Data Representation Is Important

- 1 Makes raw data **easier to interpret**
- 2 Helps in **comparing and summarizing**
- 3 Useful for **decision-making**
- 4 Helps in **identifying patterns and trends**

# Summary Table

Method	Best For	Data Type
Table	Organized presentation	Any
Bar Chart	Comparison of categories	Categorical/Discrete
Histogram	Frequency distribution	Continuous Quantitative
Pie Chart	Showing parts of a whole	Categorical
Line Graph	Showing trends over time	Time Series
Scatter Plot	Showing relationships between vars	Quantitative