Descriptive Analytics

Understanding Key Statistical Measures and Patterns

What is Descriptive Analytics?

- Descriptive analytics is the process of summarizing historical data to identify patterns.
- Common techniques include central tendency, dispersion, and shape of data distribution.
- Helps businesses understand past behavior and performance.

Measures of Central Tendency

- Mean (Average): Sum of all values divided by count.

  Formula: Mean = (Σx) / n
- Median: Middle value in an ordered list.
- Mode: Most frequently occurring value.

Mean Example

- Values: 5, 10, 15, 20, 25
- Mean = (5 + 10 + 15 + 20 + 25) / 5 = 75 / 5 = 15

Median Example

- Odd Count: 5, 10, 15 → Median = 10
- Even Count: 5, 10, 15, 20 → Median = (10 + 15)/2 = 12.5

Mode Example

- Values: 2, 4, 4, 6, 8 → Mode = 4 (occurs most often)

Measures of Dispersion

- Minimum (MIN): Smallest value in dataset.
- Maximum (MAX): Largest value in dataset.
- Range: MAX - MIN
- Standard Deviation (σ): Measure of spread around the mean.

  Formula: $\sigma = \text{sqrt}[\ \Sigma(x_i - \bar{x})^2 / n\ ]$

Standard Deviation Example

- Values: 5, 10, 15
- Mean = 10
- Variance = $[(5-10)^2 + (10-10)^2 + (15-10)^2]/3 = (25+0+25)/3 = 16.67$
- Standard Deviation = $\sqrt{16.67} \approx 4.08$

Understanding Bell Curve

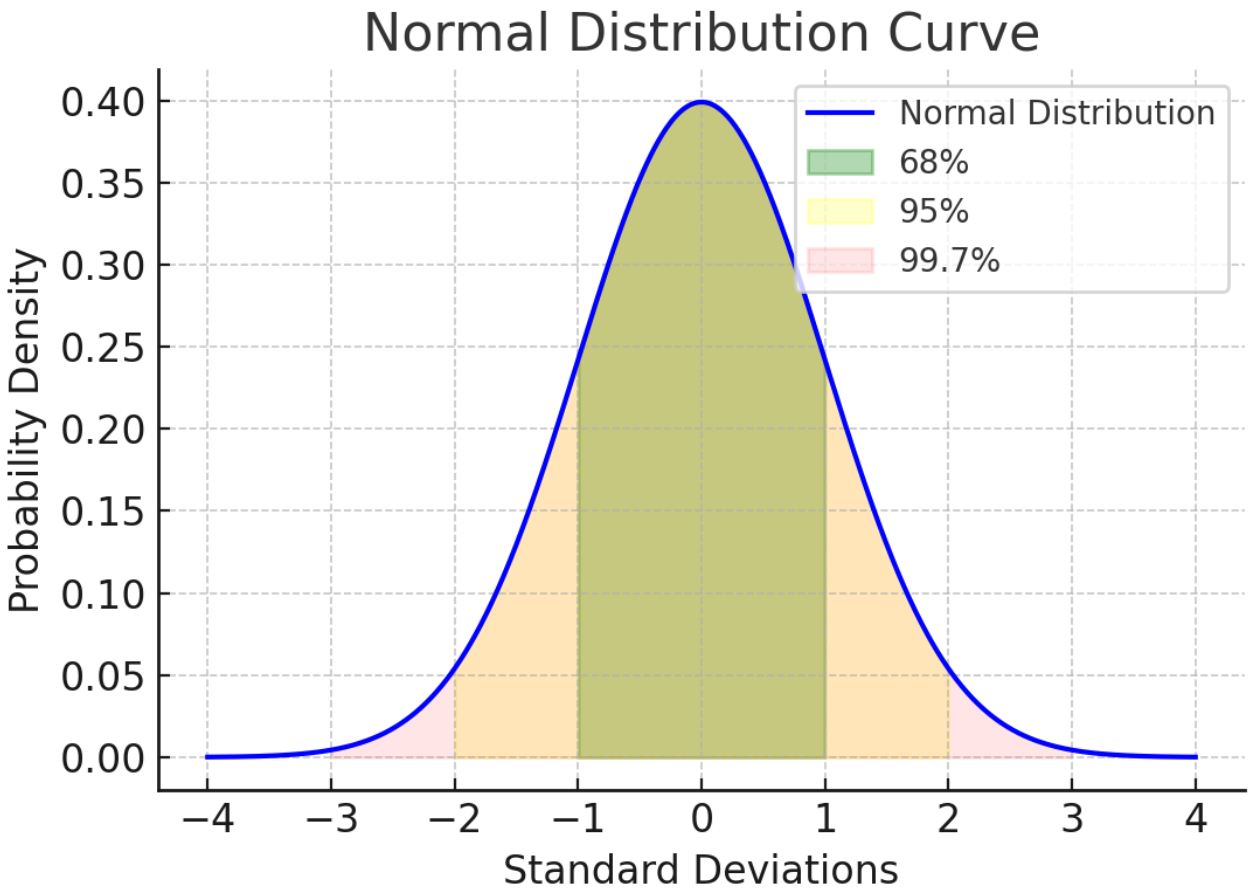
- Normal distribution is symmetric around the mean.
- 68% of data lies within 1 SD, 95% within 2 SDs, 99.7% within 3 SDs.
- Bell-shaped curve is common in naturally occurring datasets.

Bell Curve Characteristics

- Mean = Median = Mode
- Data is symmetrically distributed
- Used to identify probabilities and outliers

Normal Distribution Curve

The curve shows the spread of data around the mean with standard deviation zones.



Outliers in Data

- Outliers are extreme values that differ significantly from others.
- Can affect the mean significantly.
- Must be analyzed to decide if they are errors or valid extremes.

Range in Descriptive Analytics

- Range: Difference between the maximum and minimum values in a dataset.

  Formula: Range = MAX - MIN
- Indicates the spread of the data.
- Useful for understanding variability in small datasets.

- Example:

  Values: 10, 15, 25, 30

  Range = 30 - 10 = 20

Data Visualizations in Descriptive Analytics

- Visualizations help communicate patterns and trends clearly.
- Common visualization types include:
  - Bar Graphs – Compare quantities across categories
  - Pie Charts – Show proportions of a whole
  - Histograms – Display frequency distributions
  - Line Graphs – Show trends over time
- Choose the right chart to match your data and message.

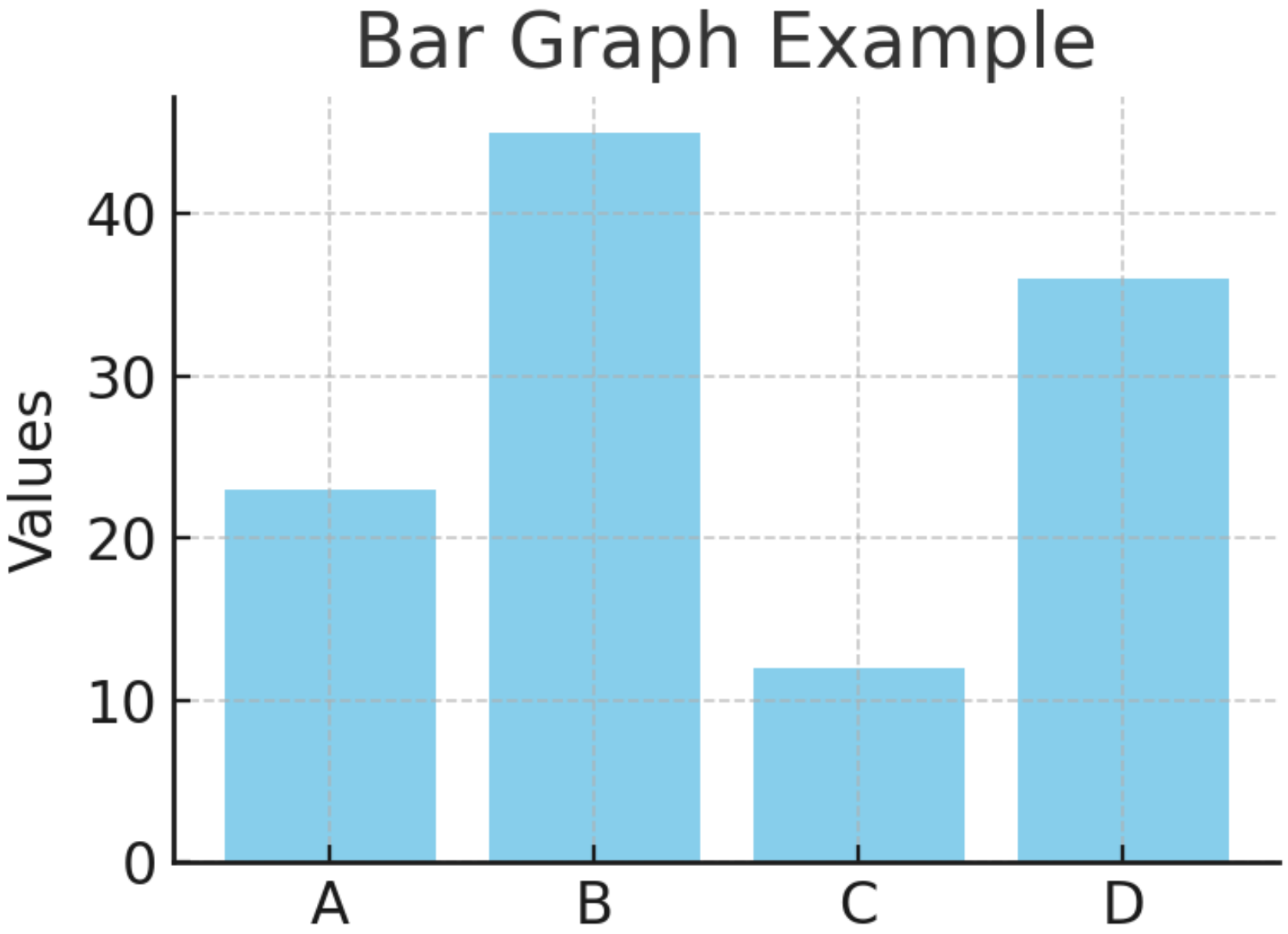
Data Visualization Techniques

Bar Graphs, Line Graphs, Pie Charts, and Histograms

Bar Graphs

- Bar graphs use rectangular bars to represent data values.
- Best for comparing quantities across discrete categories.
- Bars can be horizontal or vertical.

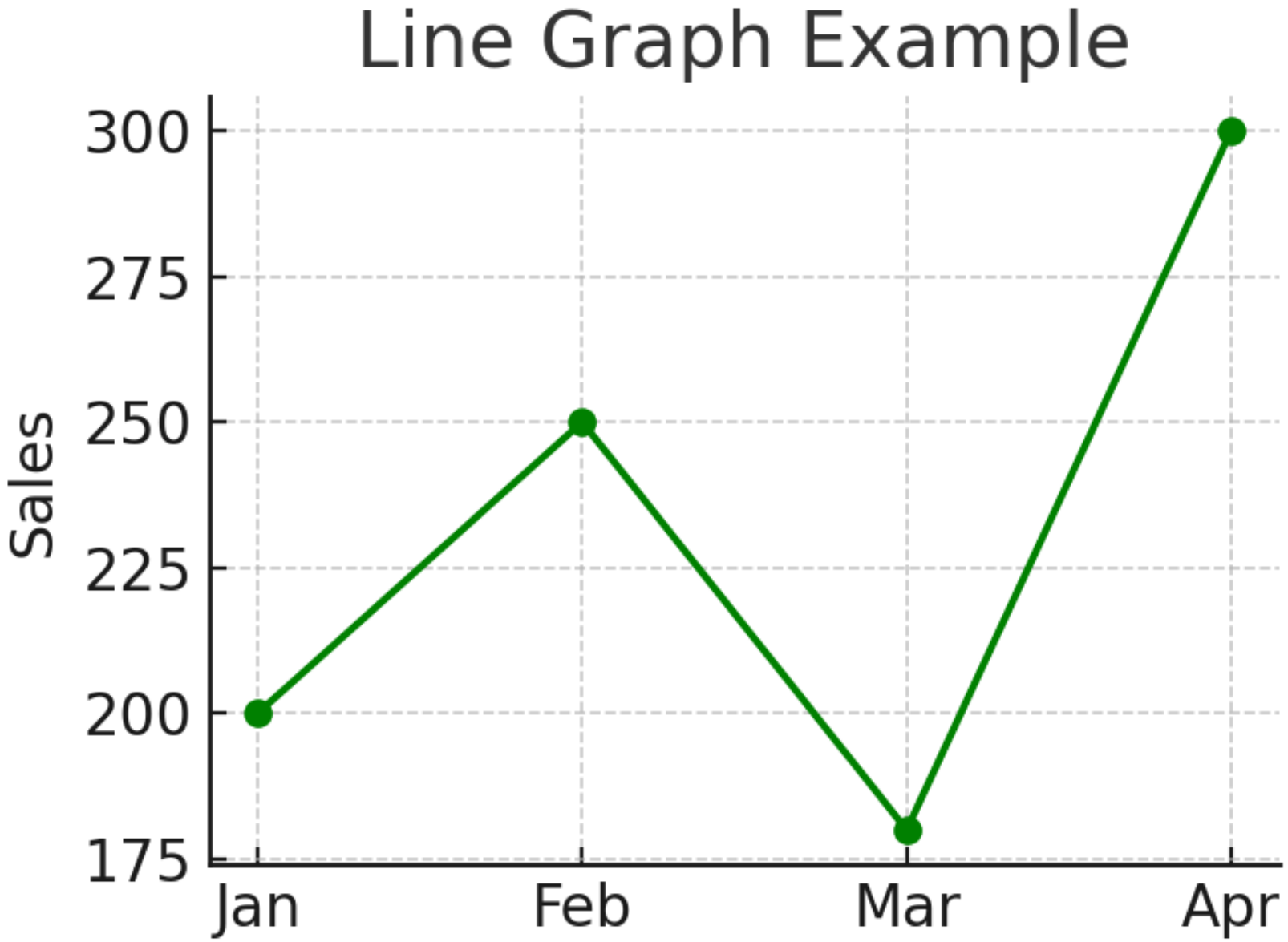
- Example use cases:
  - Sales by region
  - Survey responses by category


Bar Graph Example
Values
40
30
20
10
0
A
B
C
D

Line Graphs

- Line graphs show trends over time using connected data points.
- Ideal for continuous data like temperature, revenue, or growth.
- Useful for highlighting changes and trends.

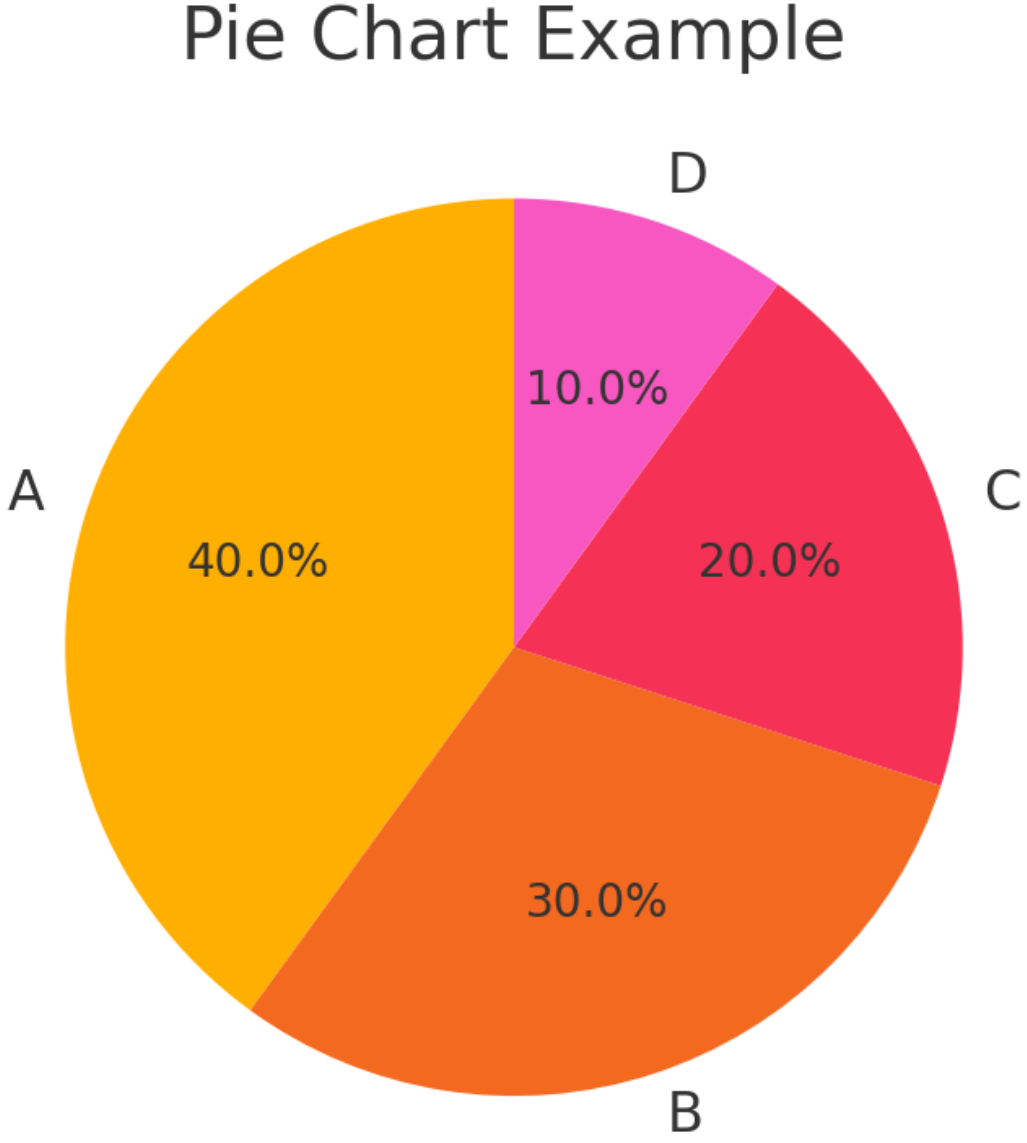
- Example use cases:
  - Monthly sales growth
  - Website traffic over weeks


Line Graph Example
Sales
300
275
250
225
200
175
Jan
Feb
Mar
Apr

Pie Charts

- Pie charts show proportions as slices of a whole.
- Each slice represents a percentage of the total.
- Best used when there are fewer categories.

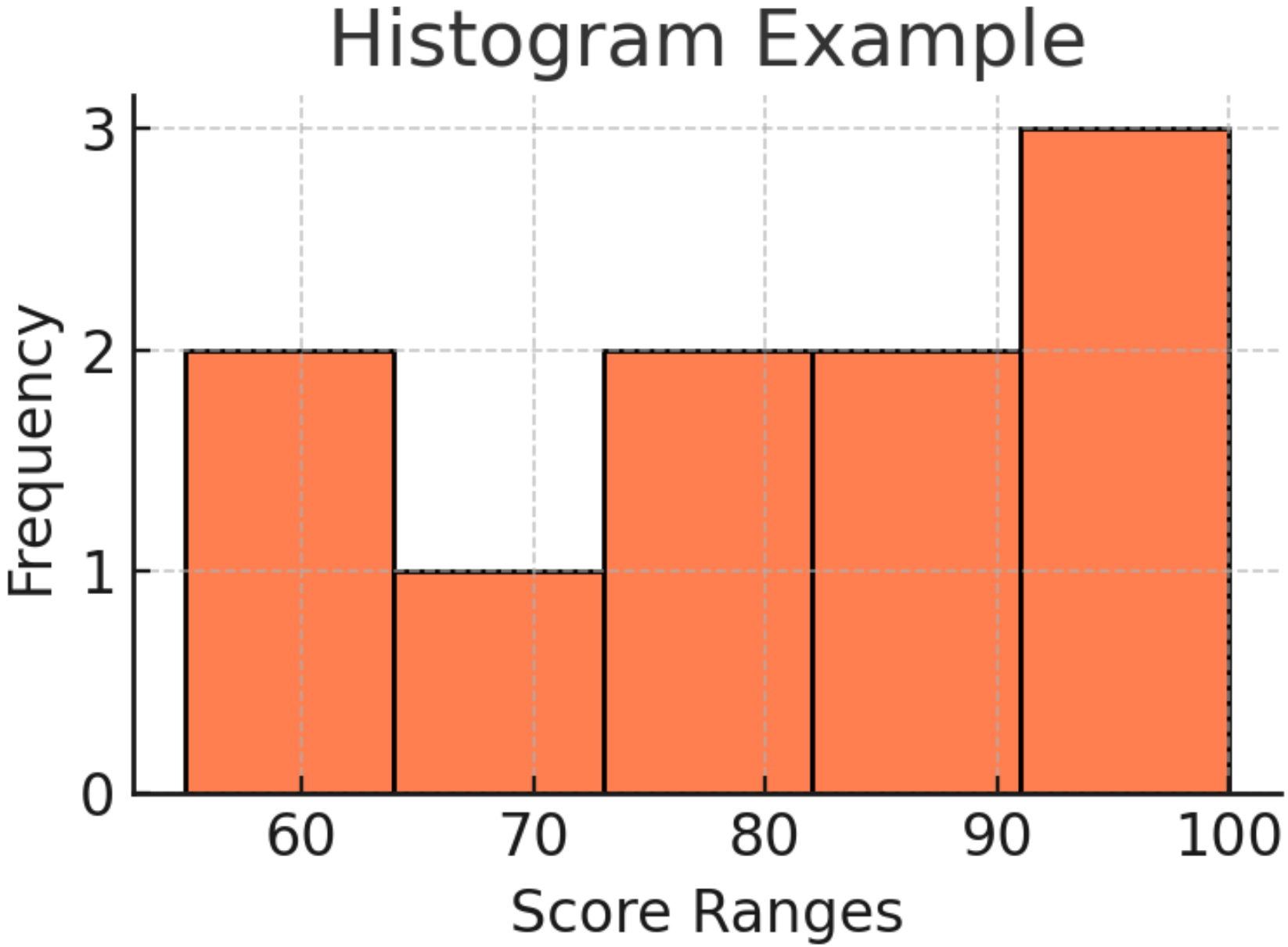
- Example use cases:
  - Market share by company
  - Budget breakdown by department


Pie Chart Example
D
10.0%
A
C
40.0%
20.0%
30.0%
B

Histograms

- Histograms represent frequency distributions of continuous data.
- Data is grouped into bins or intervals.
- Helps visualize the shape of the data distribution.

- Example use cases:
  - Distribution of exam scores
  - Frequency of transaction amounts


Histogram Example
Frequency
0
1
2
3
60
70
80
90
100
Score Ranges

Understanding Pivot Tables

Powerful Data Summarization in Excel and Google Sheets

What is a Pivot Table?

- A pivot table is a tool that allows you to summarize, analyze, explore, and present data.
- It automatically sorts, counts, and totals the data stored in one table.
- Ideal for large datasets where quick insights are needed.

Why Use Pivot Tables?

- Quickly summarize large data sets
- Identify trends and patterns
- Create interactive reports
- Easily compare data between categories

Parts of a Pivot Table

- Rows: Categories to display vertically
- Columns: Categories to display horizontally
- Values: Data to aggregate (sum, average, count)
- Filters: Criteria to limit what data is included

Creating a Pivot Table (Google Sheets)

1. Select your dataset
2. Go to Data > Pivot table
3. Choose destination sheet
4. Use editor to set Rows, Columns, Values, and Filters

Common Aggregations

- SUM: Total values in a field
- AVERAGE: Mean of values
- COUNT: Number of entries
- MAX / MIN: Highest or lowest values

Example Use Case

- Dataset: Sales transactions by product and region
- Pivot Table Output:
  - Rows: Product
  - Columns: Region
  - Values: Sum of Sales
- Insight: Identify top-performing products by region

Best Practices

- Clean your data before creating a pivot table
- Use descriptive headers
- Avoid merging cells in your raw data
- Use slicers/filters to drill down into details

Descriptive Statistics with Google Sheets

Analyzing and Summarizing Data Easily in the Cloud

Why Use Google Sheets?

- Free and cloud-based
- Accessible from any device with a browser
- Built-in statistical functions
- Supports charts and pivot tables
- Real-time collaboration

Key Functions in Google Sheets

- =AVERAGE(range) – Calculates the mean
- =MEDIAN(range) – Finds the middle value
- =MODE(range) – Finds the most frequent value
- =STDEV(range) – Calculates standard deviation
- =MIN(range), =MAX(range) – Find min and max

Example Dataset

- Columns: Product, Region, Sales, CustomerAge
- Apply functions to the Sales column to compute:
  - =AVERAGE(B2:B101)
  - =STDEV(B2:B101)
  - =MAX(B2:B101) - MIN(B2:B101) for range

Creating Charts

- Select your data
- Go to Insert > Chart
- Choose appropriate chart type:
  - Bar chart for comparisons
  - Line chart for trends
  - Histogram for frequency
  - Pie chart for proportions

Using Pivot Tables

- Go to Data > Pivot table
- Drag fields into Rows, Columns, and Values
- Use SUM, AVERAGE, or COUNT for aggregation
- Example: Total Sales by Region

Best Practices

- Always clean your data first
- Label all columns clearly
- Use conditional formatting to highlight outliers
- Save copies before making major edits
- Use filters to explore subsets of data