

# **DESCRIPTIVE STATISTICS IN QUANTITATIVE DATA ANALYSIS**

29 October 2024

# QUANTITATIVE DATA ANALYSIS

Quantitative data analysis focuses on analyzing numerical data to identify patterns, relationships, and trends.

---

- Quantitative data analysis deals with numerical data.
- Helps identify patterns, relationships, and trends.
- Used to summarize large datasets.

---

Data Analysis methods

---

# QUANTITATIVE DATA ANALYSIS

**Descriptive statistics are brief informational coefficients that summarize a given data set, which can be either a representation of the entire population or a sample of a population.**

---

- Descriptive statistics summarize and organize data.
- Focus on measures of central tendency and variability.
- Does not infer or predict, but describes the data.

---

Data Analysis methods

---

# CATEGORIES

**Descriptive statistics are typically divided into three categories:**

---

- **Measures of Central Tendency (mean, median, and mode)** that describe the center of the data.
- **Measures of Variability (standard deviation, variance, range)** that describe the spread of the data.
- **Frequency Distribution** that describes how often values appear in the data.

---

Data Analysis methods

---

# MEASURES OF CENTRAL TENDENCY

**Measures of central tendency describe the center of the data and indicate what is typical or common within the dataset.**

---

- There are three primary measures:
- mean
- median
- mode

---

Data Analysis methods

---

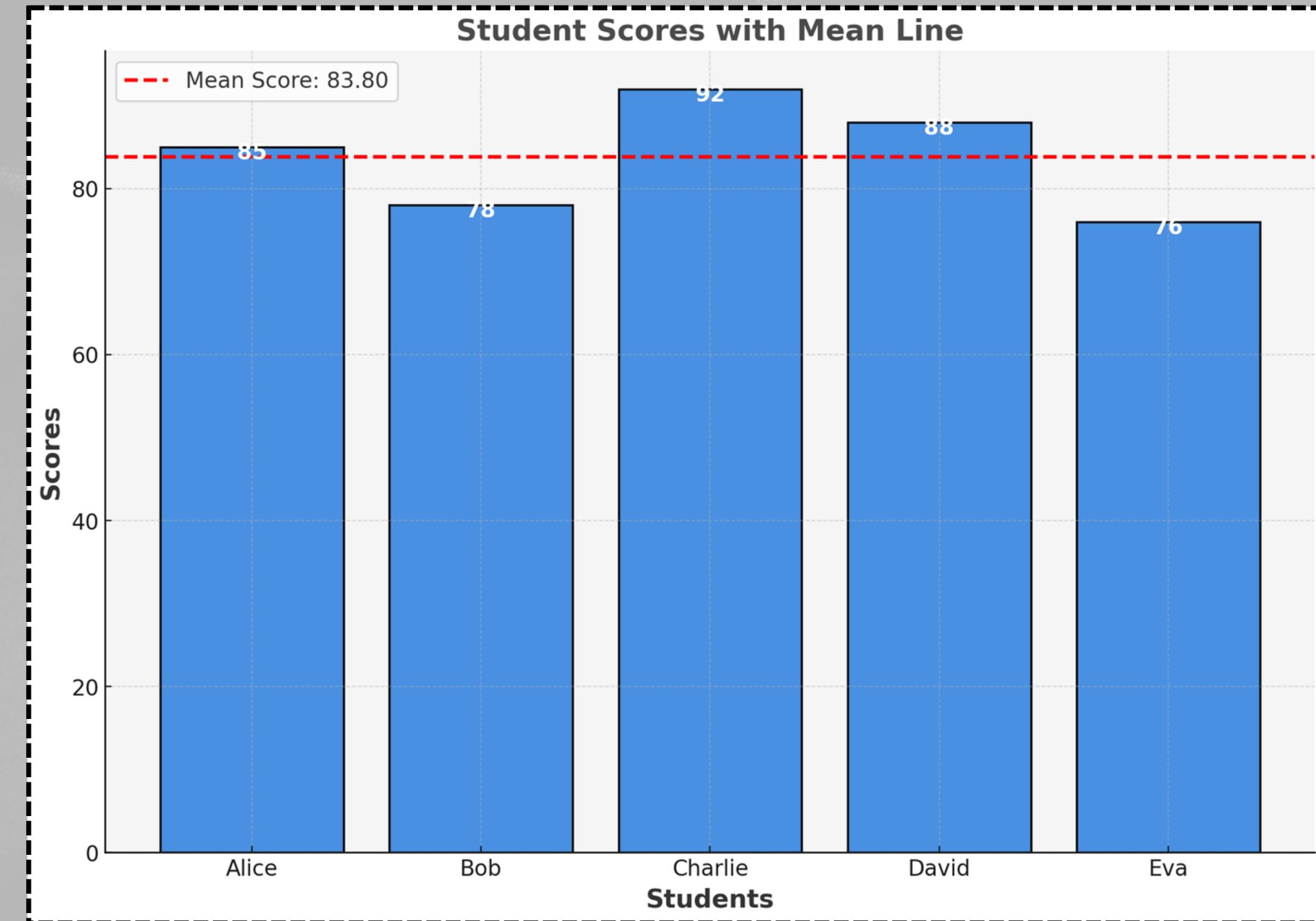
# CENTRAL TENDENCY: MEAN

The average value of a dataset.

---

**Calculation:** Sum of all values divided by the number of values.

**Example:** Mean exam score from 5 students:  $(70 + 80 + 90 + 60 + 85) / 5 = 77$ .



Data Analysis methods

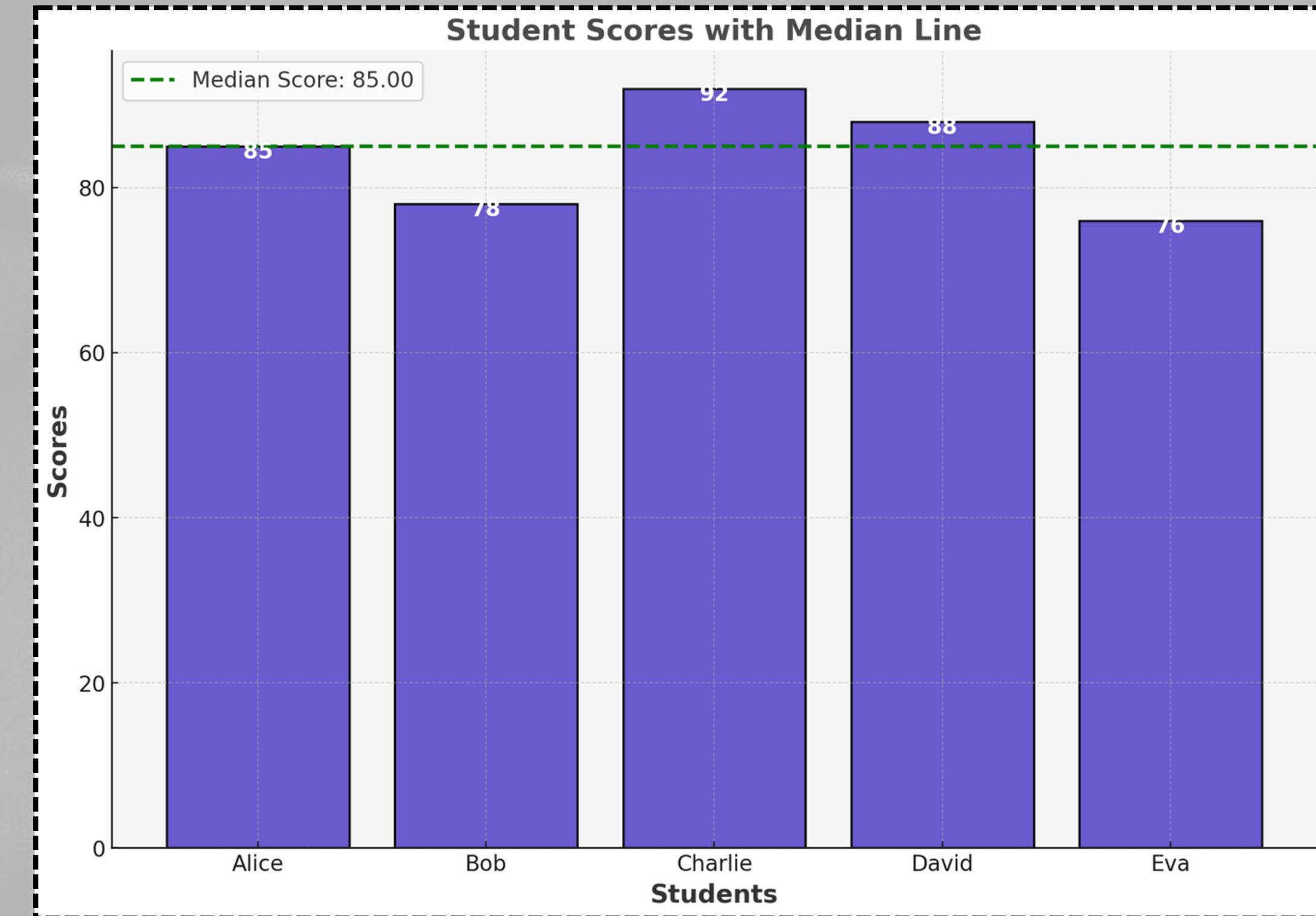
# CENTRAL TENDENCY: MEDIAN

**The middle value when data is ordered.**

---

**Use:** Helps identify the midpoint of the data.

**Example:** Exam scores (sorted): 60, 70, 80, 85, 90. Median = 80.



---

Data Analysis methods

---

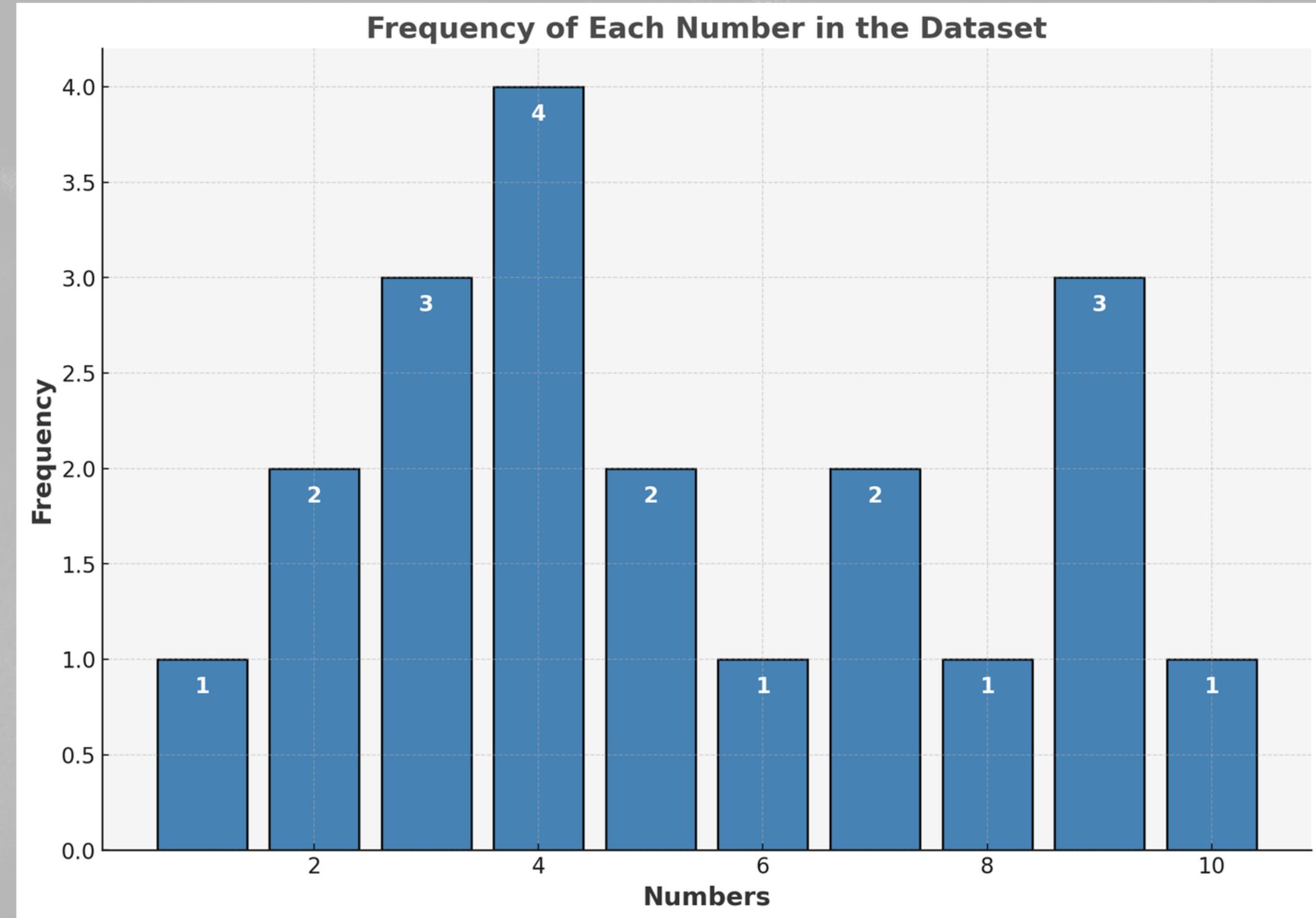
# CENTRAL TENDENCY: MODE

**The most frequent value in a dataset.**

---

**Use:** Helps understand common or recurring data points.

**Example:** Shoe sizes: 6, 7, 7, 8, 9. Mode = 7.



---

Data Analysis methods

---

# MEASURES OF VARIABILITY (SPREAD)

Measures of variability give insight into the dispersion of the dataset, showing whether the data points are closely clustered or widely scattered.

---

The key measures of variability include:

- Range
- Variance
- Standard Deviation

---

Data Analysis methods

---

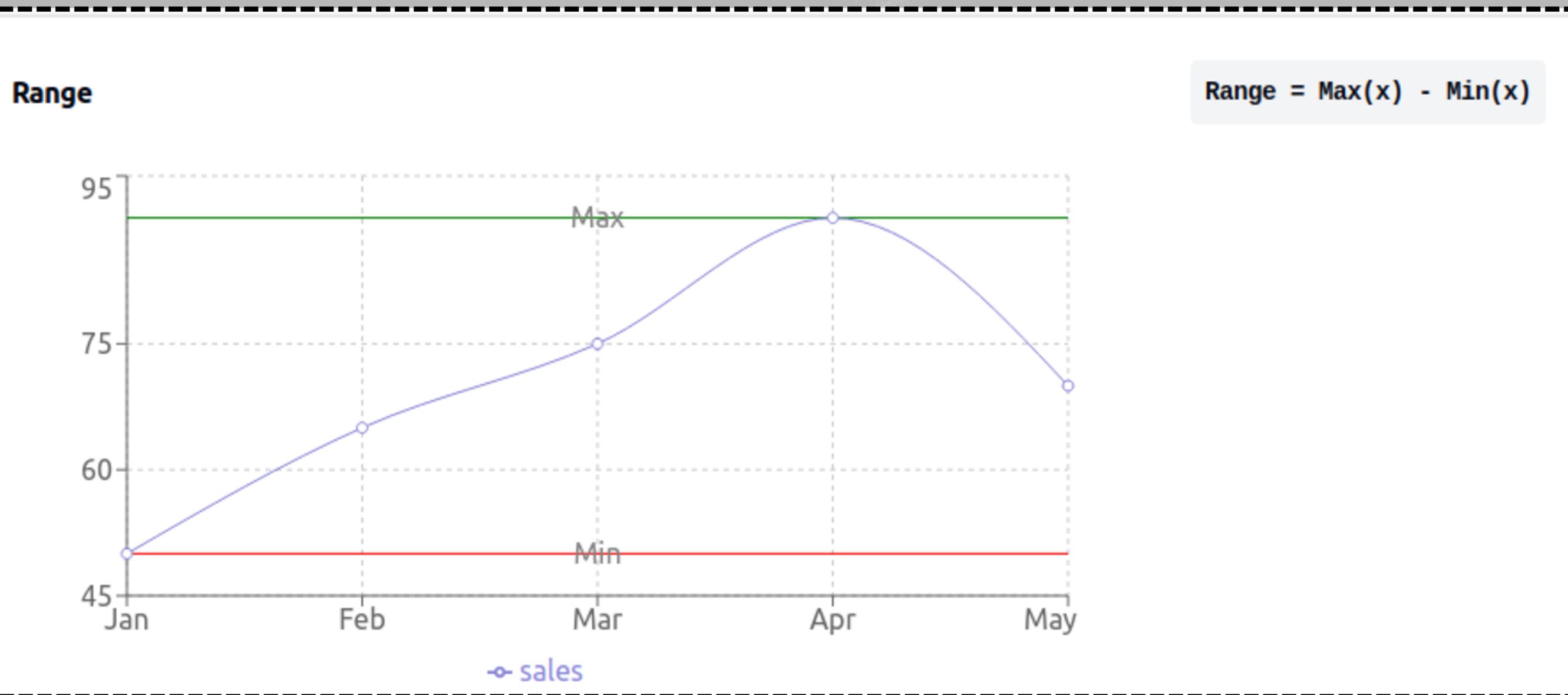
# VARIABILITY: RANGE

**The difference between the highest and lowest values.**

---

**Calculation:** Maximum value - Minimum value.

**Example:** Sales data: \$50, \$70, \$90. Range =  $90 - 50 = \$40$ .



Data Analysis methods

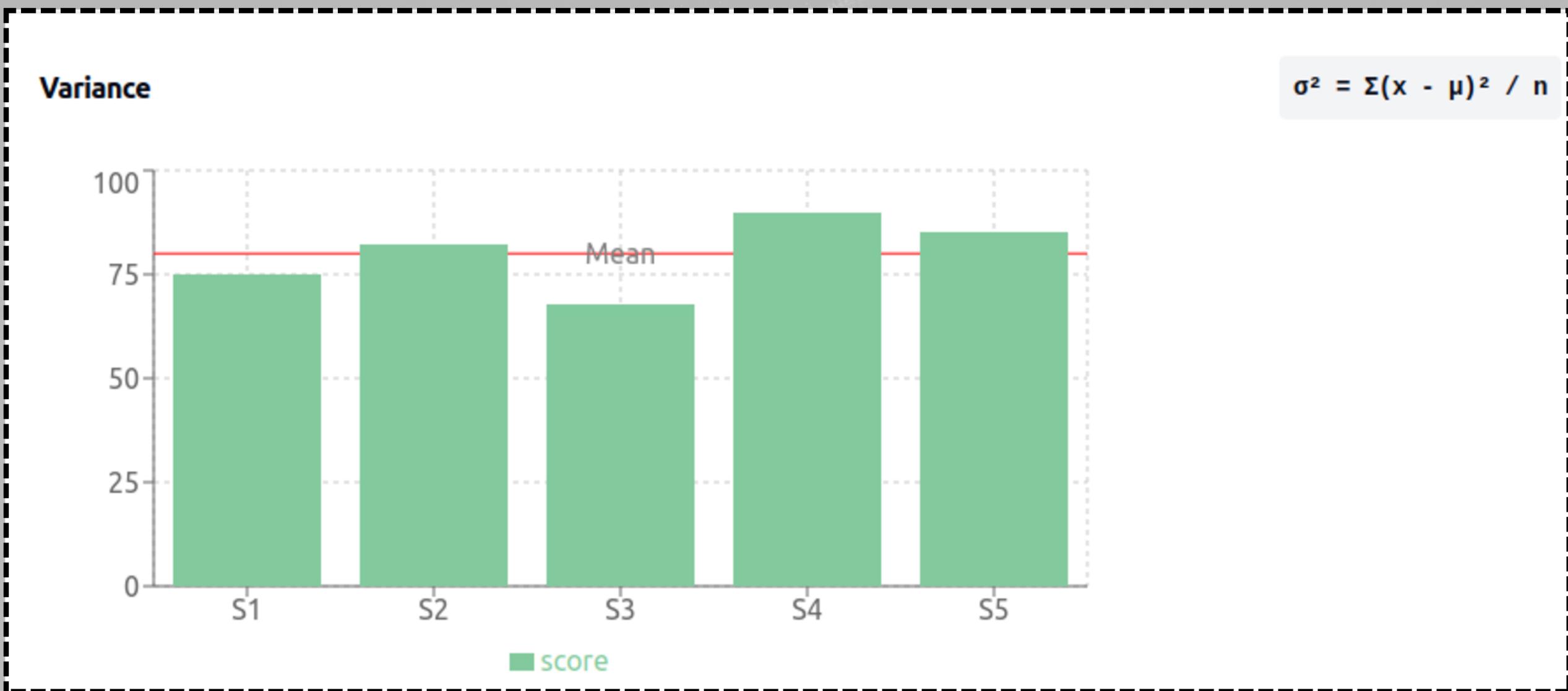
# VARIABILITY: VARIANCE

**Measures how far each data point is from the mean.**

---

**Use:** Understanding how spread out the data is.

**Example:** Variance of exam scores = 100.



---

Data Analysis methods

---

# VARIABILITY: STANDARD DEVIATION

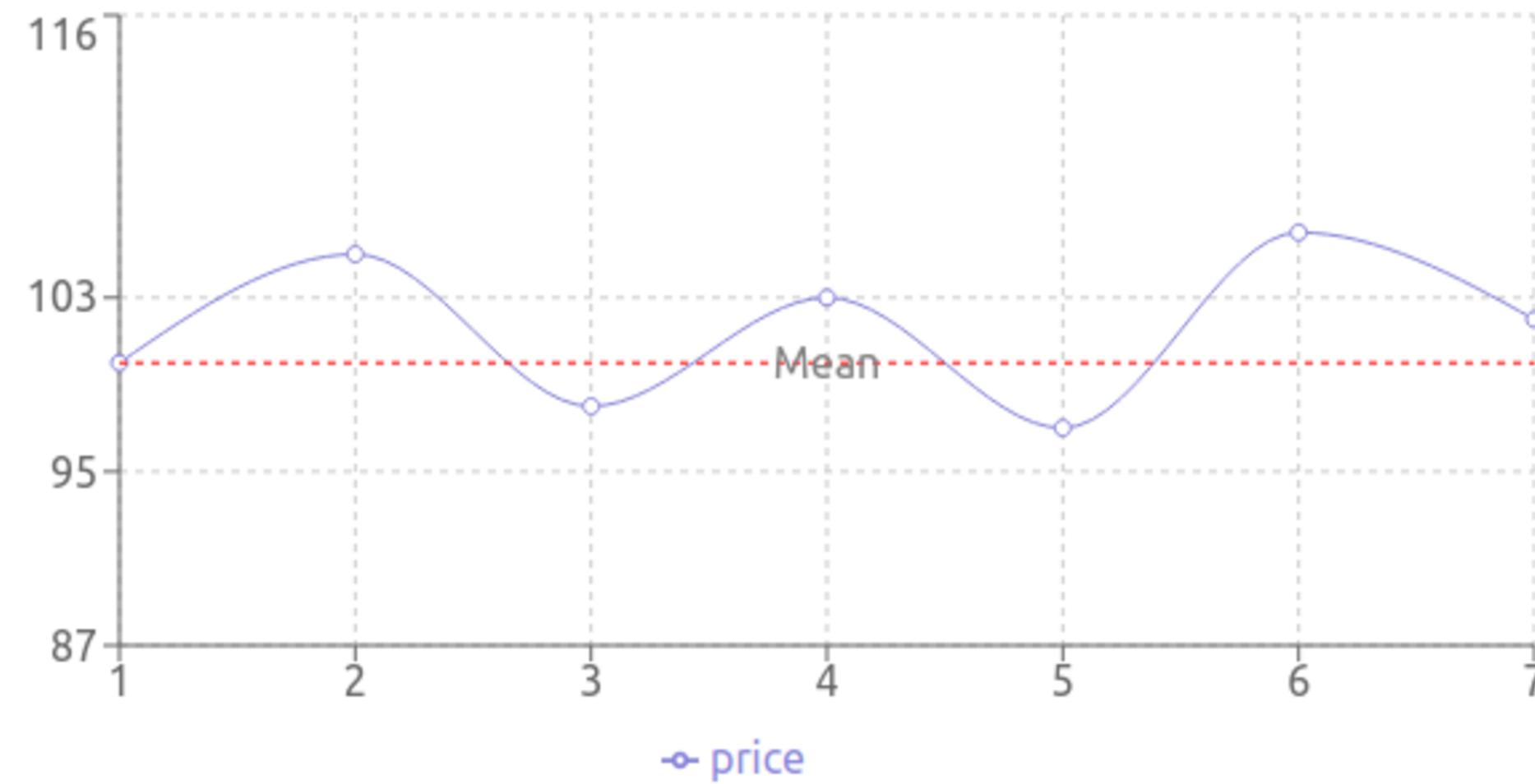
**The square root of the variance.**

---

**Use:** Measures how closely data points cluster around the mean.  
**Example:** Standard deviation of daily stock prices = 5.

### Standard Deviation

$$\sigma = \sqrt{(\Sigma(x - \mu)^2 / n)}$$



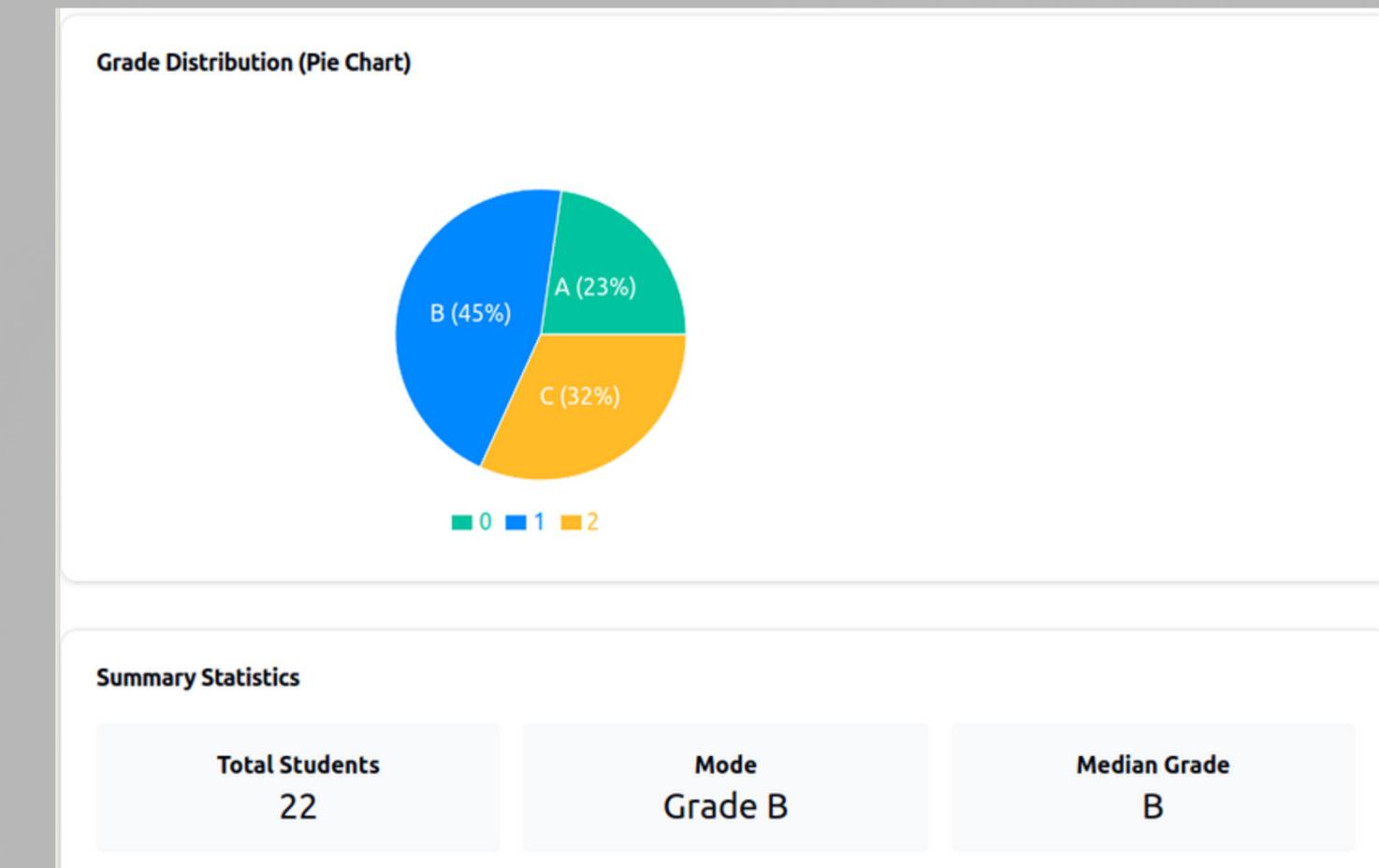
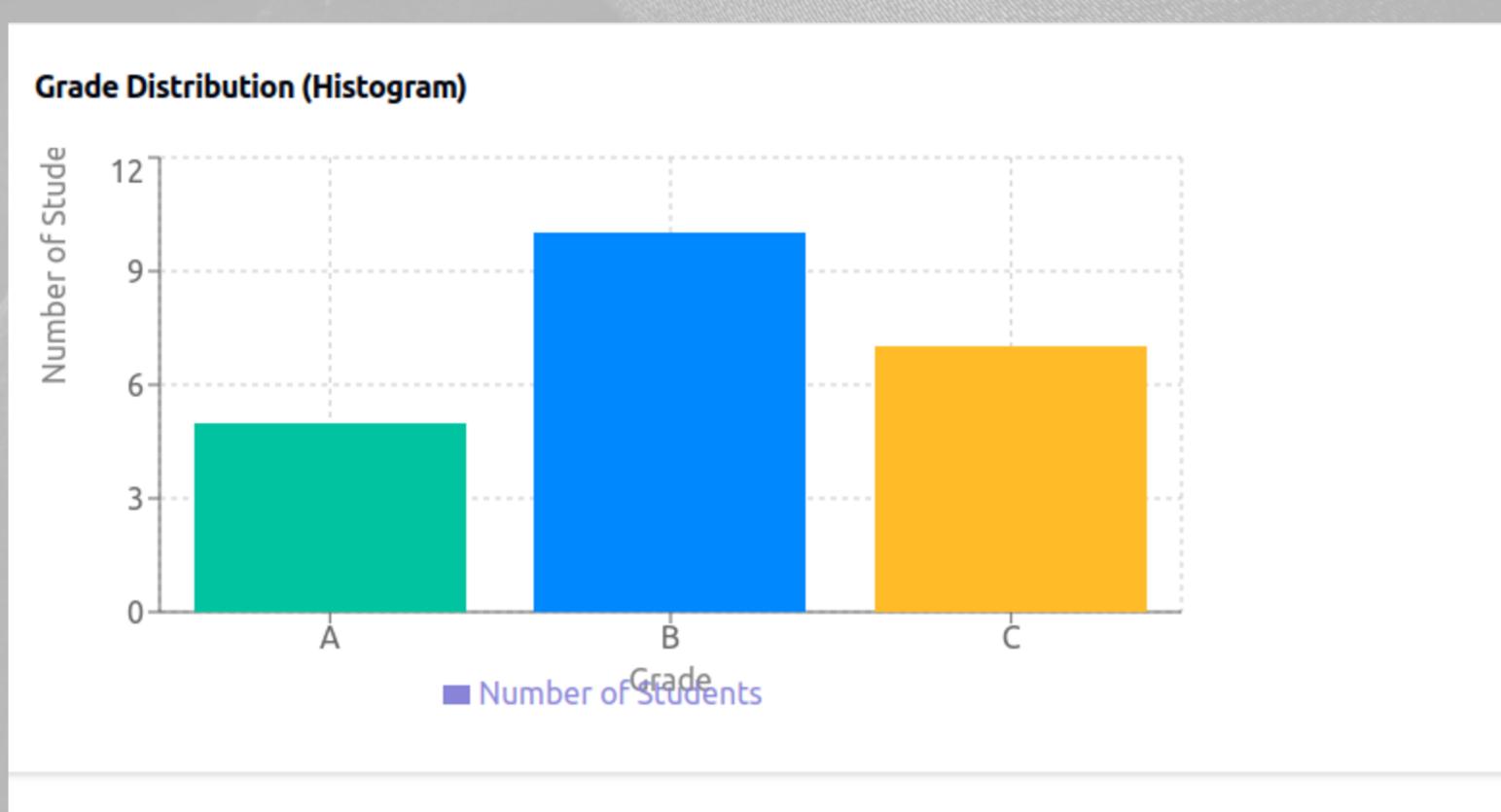
Data Analysis methods

# FREQUENCY DISTRIBUTION

**Shows how often each value appears in a dataset.**

---

**Use:** Useful for identifying trends or common values.  
**Example:** Student grades: A (5 students), B (10 students), C (7 students).



---

Data Analysis methods

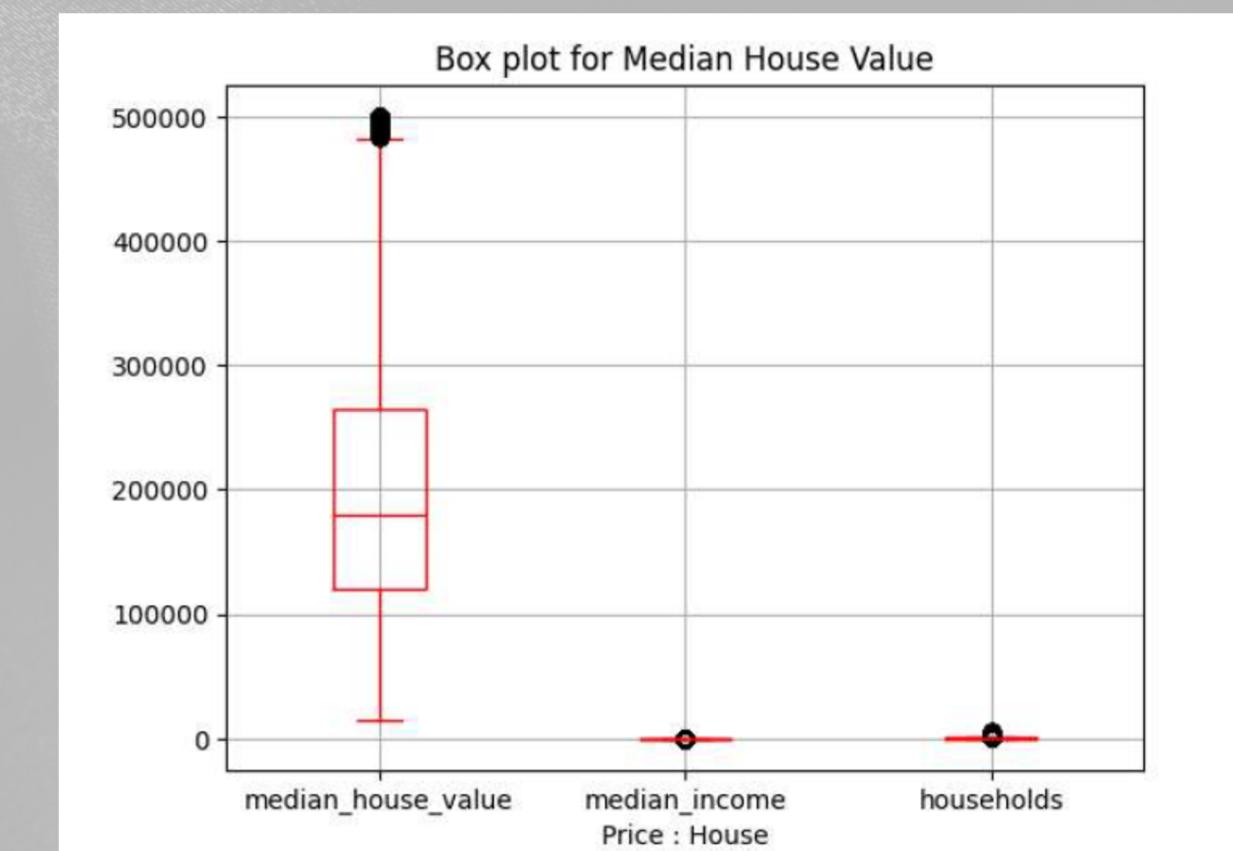
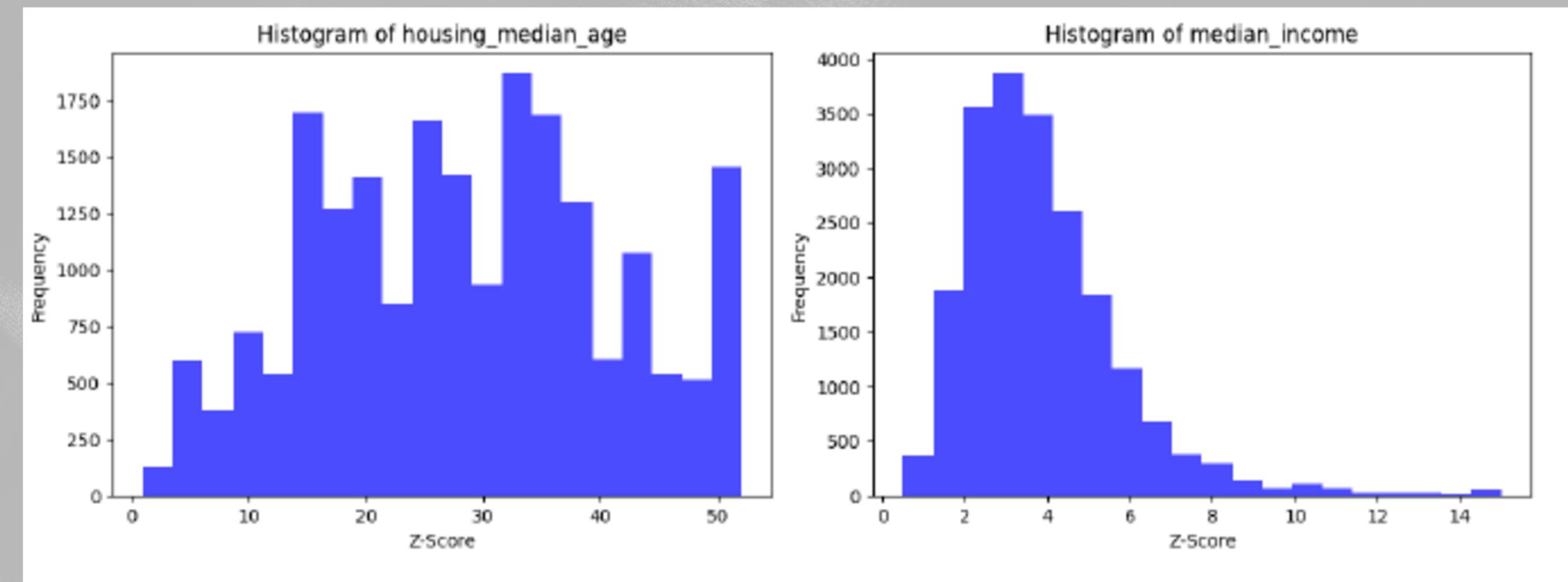
---

# VISUAL REPRESENTATIONS IN DESCRIPTIVE STATISTICS

**Histograms:** A histogram displays the frequency of data points within specified ranges, using bars to represent the count of data points in each range.

**Box Plots (Box-and-Whisker Plots):** Box plots provide a graphical representation of the distribution of data, highlighting the median, quartiles, and any potential outliers. This makes box plots an excellent tool for identifying whether the data is skewed and how spread out it is.

**Bar Charts:** Bar charts are used to compare the frequency of different categories. They are especially useful when dealing with categorical data, such as comparing the sales performance of different products over a given period.



---

Data Analysis methods

---

# TOOLS USED FOR DESCRIPTIVE STATISTICS

**Python** : Data analysis and manipulation.

**Excel**: Basic data analysis and visualization.

**SPSS**: Widely used for descriptive analysis in social sciences.

# CONCLUSION

- Descriptive statistics are essential for summarizing large datasets.
- Provides key insights without making predictions.
- Visualizations enhance understanding of data patterns.
- A foundation for advanced data analysis like inferential statistics.

**Call to Action:** Understanding descriptive statistics is critical for any data-driven decision-making process.