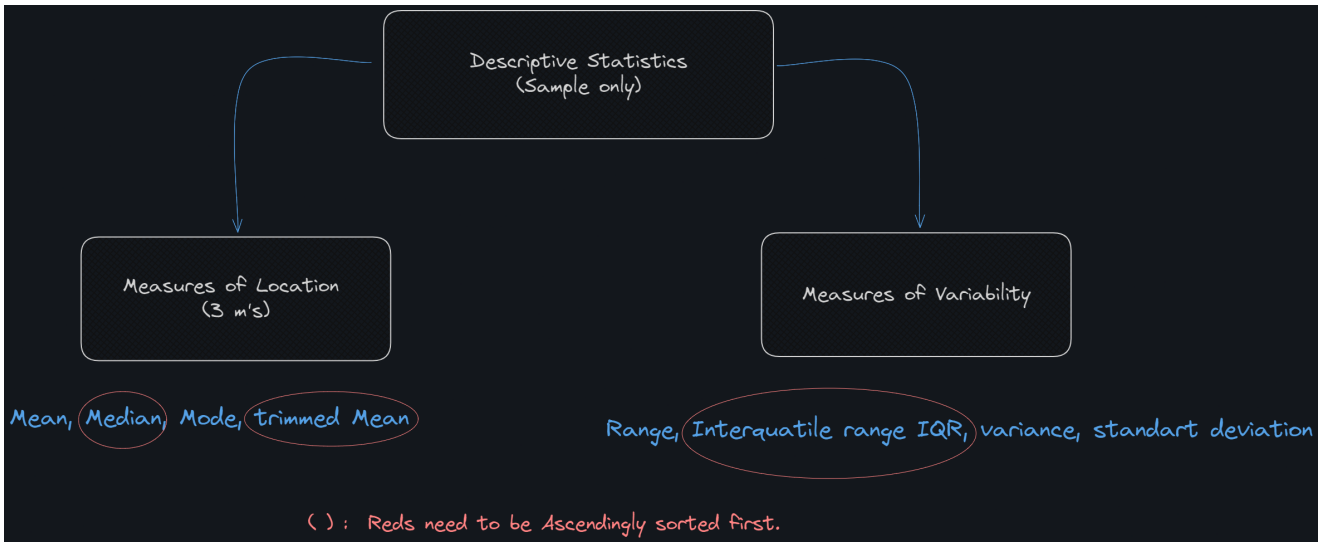
Descriptive Statistics



Measures of Location

1. Mean (Average):

- **Formula:**

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- **Description:** The sum of all values in the dataset divided by the number of values.
- **Sensitivity:** Sensitive to outliers, meaning extreme values can significantly affect the mean.

2. Median:

- **Formula**

$$\tilde{x} = \frac{x(n+1)}{2}$$

- **Description:** The middle value when the data is arranged in ascending or descending order.
- **Calculation:**
  - If n (number of values) is odd: Median = (n + 1)/2 th value.
  - If n is even: Median = the average of the (n/2)th and (n/2 + 1)th values.
- **Sensitivity:** Less sensitive to outliers compared to the mean.

3. Mode:

- **Description:** The value that appears most frequently in the dataset.
- **Calculation:** Identify the value with the highest frequency.
- **Uniqueness:** A dataset can have multiple modes (bimodal, multimodal), or no mode if all values appear equally often.

4. Trimmed Mean:

- **Description:** Similar to the mean, but excludes a specified percentage of extreme values from both ends of the data.
- **Calculation:**
    1. Order the data.
    2. Exclude a pre-defined percentage of values from each tail (e.g., 10%).
    3. Calculate the mean of the remaining values.
- **Purpose:** Reduces the impact of outliers on the central tendency measure.

Additional Important Information:

- **Understanding the data distribution:** Knowing the shape of the data distribution (symmetrical, skewed) is crucial when choosing the appropriate measure.
- **Outliers:** The presence of outliers can significantly impact the mean but not necessarily the median or mode.
- **Context matters:** The best measure depends on the specific context and research question. For example, the median might be preferred for skewed distributions or when dealing with ordinal data (data with a rank but not necessarily equal intervals).