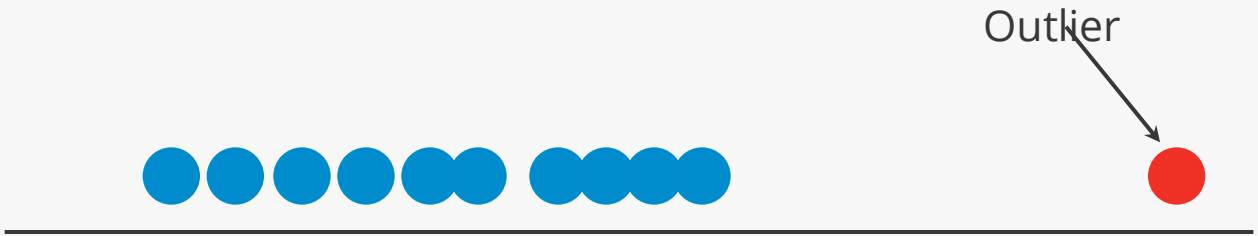
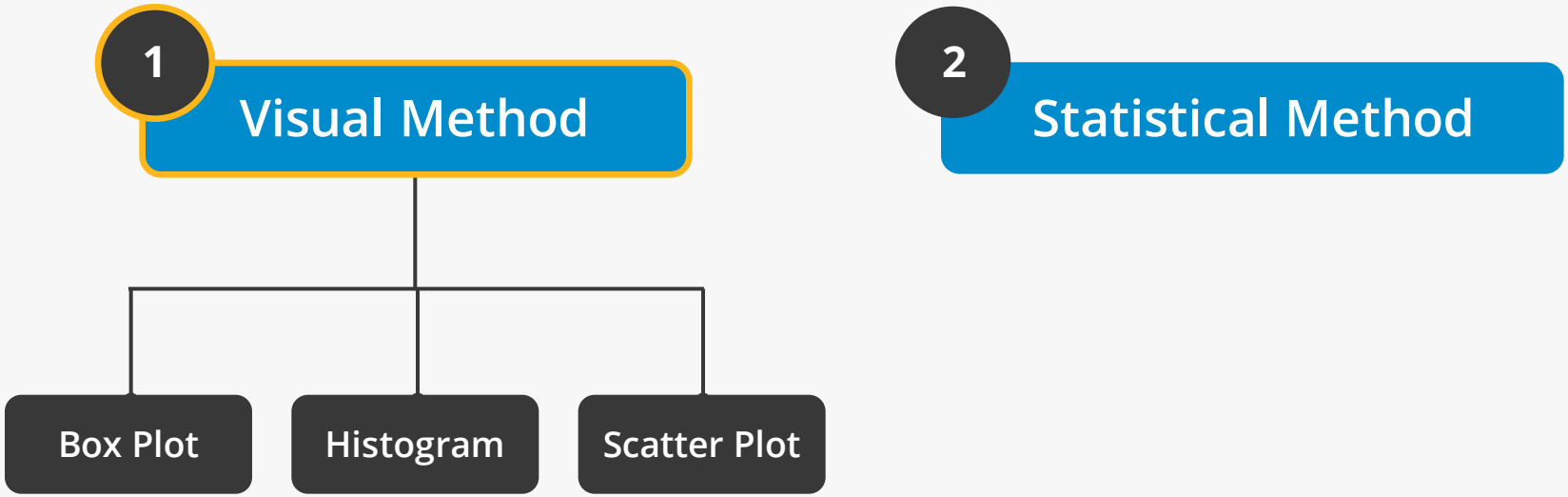
Outliers



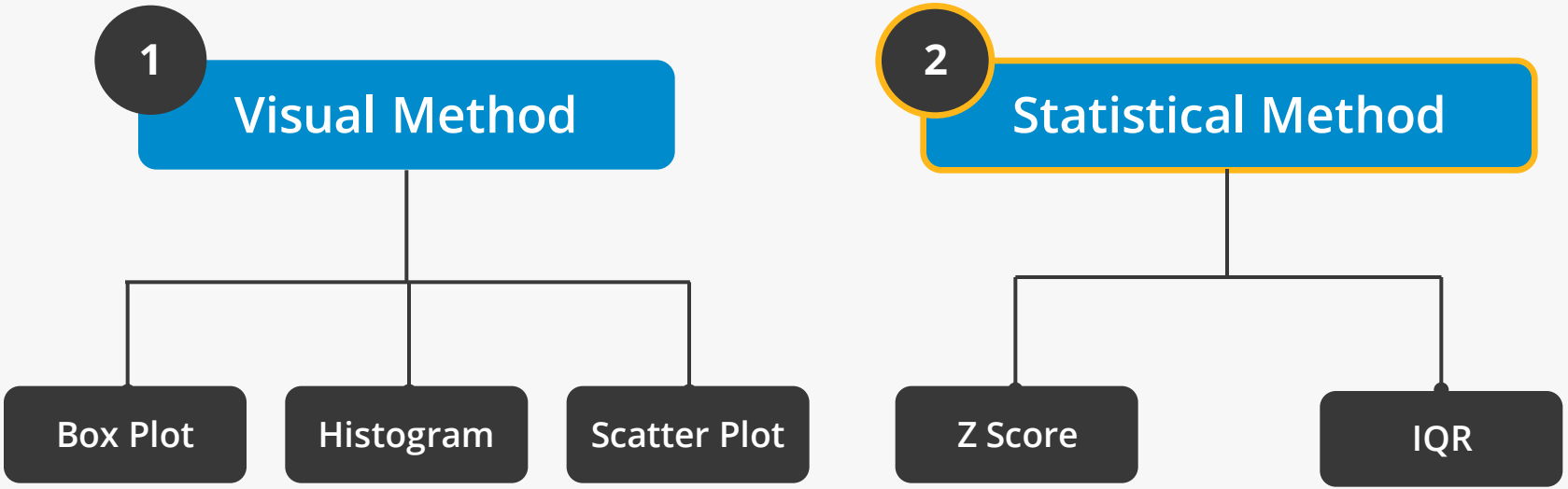
Ways to detect Outliers

1. Visual Method
2. Statistical Method

Ways to detect Outliers



Ways to detect Outliers

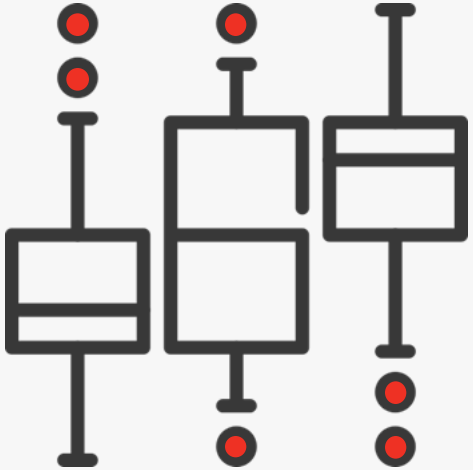


Ways to detect Outliers

1 Visual Methods

Box plots visually show the range of the data, with outliers plotted as individual points outside the whiskers.

Box Plot



Ways to detect Outliers

1 **Visual Methods**

A sudden spike or dip in a histogram can indicate outliers.

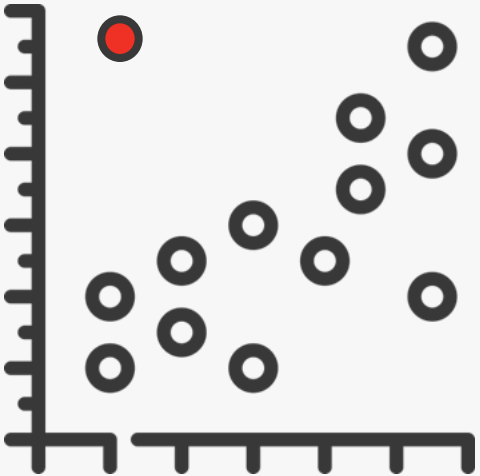
Histogram

Ways to detect Outliers

1 Visual Methods

Scatter plots can help visualize outliers as points that fall far from the main cluster of data.

Scatter Plot

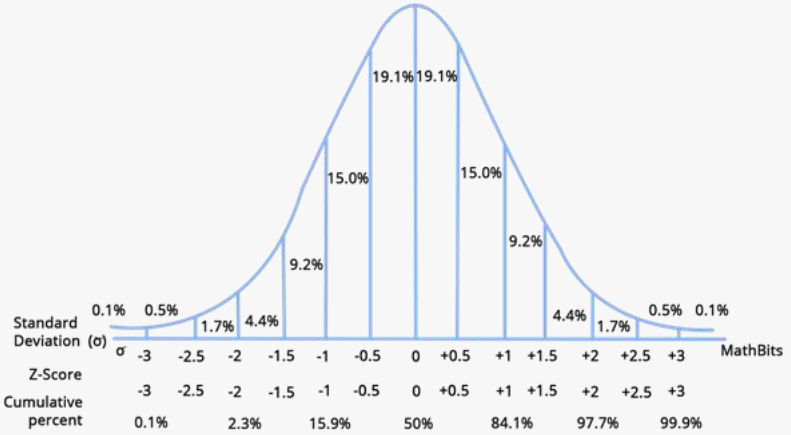


Ways to detect Outliers

2 Statistical Methods

Any datapoint which is greater than 3σ or lesser than -3σ can be considered as an outlier.

Z - Score



Ways to detect Outliers

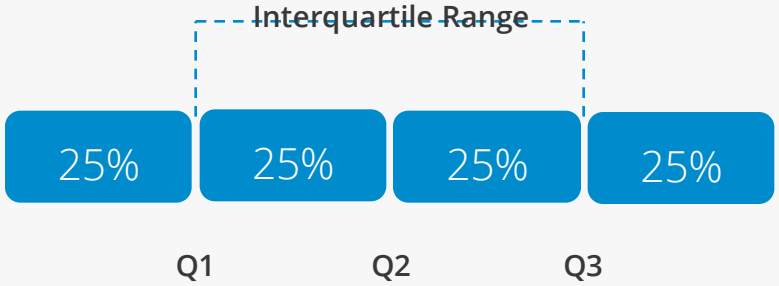
2 Statistical Methods

Data points < Q1 - 1.5 (IQR)

or

Data points > Q3 + 1.5 (IQR)

Interquartile Range (IQR)



Outlier Treatment

Outlier Detection and Treatment

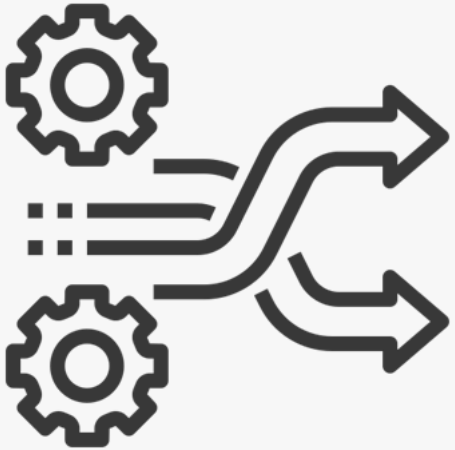
Removing the outliers

Though this practice is not recommended, it involves removing the outliers from the dataset.

Outlier Detection and Treatment

Transforming Values

Apply a mathematical function to the variable for better distribution and to reduce effect of outliers



Outlier Detection and Treatment

Quantile Based Flooring and Capping

Setting a minimum and maximum threshold and replacing all outliers with the threshold value

Outlier Detection and Treatment

Mean / Median Imputation

Replace the outlier values with the mean or median of the rest of the data

Outlier Treatment

- Domain knowledge
- Validity of the outlier
- Amount of data