LIVEWIRE
FOR LIVE CAREERS

BUSINESS ANALYTICS

Basic Statistics

- Measures of location or central tendency
- Measures of spread or Dispersion
- Frequencies
- Graphical methods or Data Visualisation

Measures of location or central tendency

- **MEAN**
- **MEDIAN**
- **MODE**

Measures of Spread or Dispersion

RANGE – Difference between highest and lowest value

Quartiles – Describes the spread of a data set by breaking the data set into quarters (Divided by 2)

- **Inter-quartile Range -** Divided by 4 (ie. 25%, 50%, 75% and 100%)
- **Inter-turtle Range -** Divided by 3
- **Inter-quantile Range -** Divided by 5

Variance

- Variance is an average of squared deviations about the mean

$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

Note: To avoid getting Zero, the deviation values are squared before they are added up.

Standard deviation

- Standard deviation is the squared root of variance:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(\bar{x} - x_i)^2}{n}}$$

Note: If the data points are too far from the mean, there is higher deviation within the data set.

Frequency

- Frequency of an event
- Represented as “n” number of times the event occurred in an experiment
- Frequencies are often graphically represented in histograms

Number of marks	Tally marks	Frequency
1	𝍸 II	7
2	𝍸	5
3	𝍸 I	6
4	𝍸	5
5	III	3
Total		**26**

Data Visualization

- Pictorial representation of data
- Helps to understand data with visual aid

Type of Data Visualization

Pie Chart

- Proportion and Percentage between categories

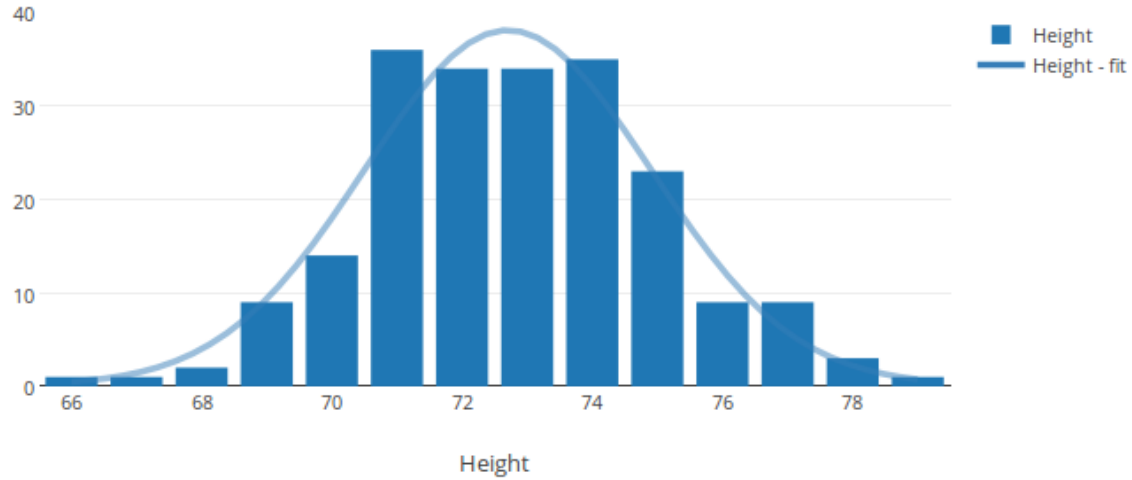
Bar Chart

- How many in each category
- Compare categorical data

Type of Data Visualization

Histogram

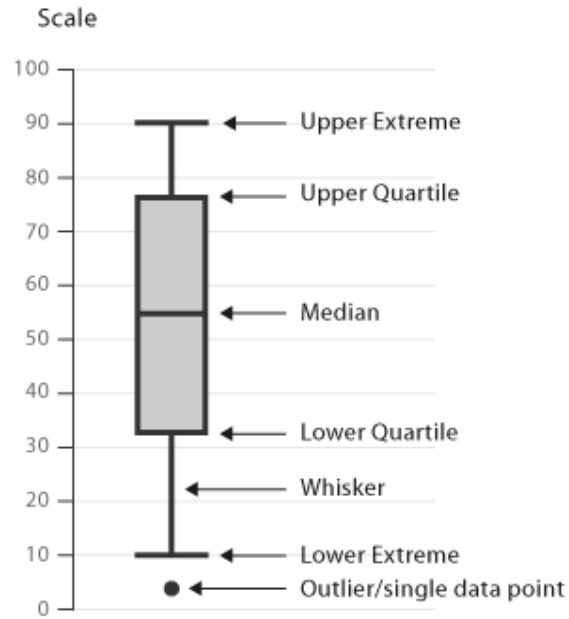
- Distribution of data over continuous interval
- Probability distribution



Source: nhl_draft_2013_@thejustinfisher

Box Plot

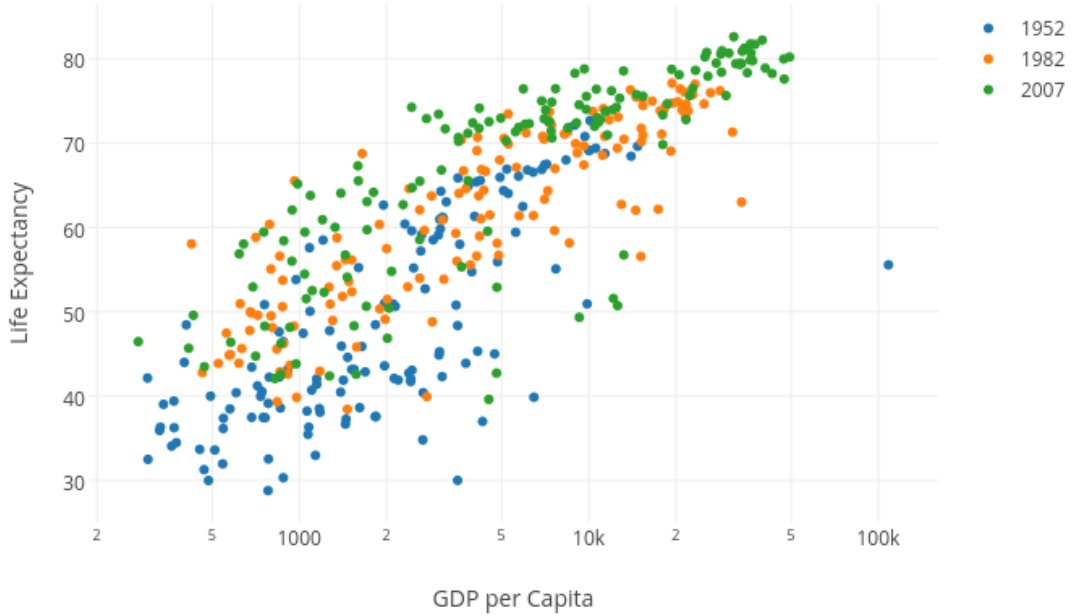
- Minimum and Maximum
- Median and Mean
- Quartile
- To know outlier



Type of Data Visualization

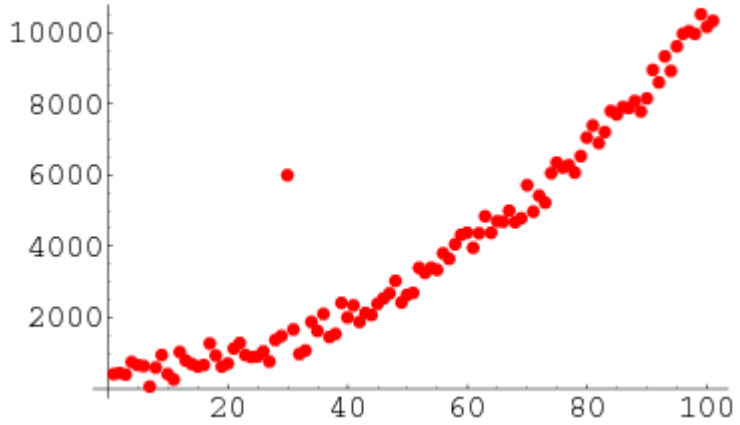
Scatter Plot

- To know relationship between variables



OUTLIER TREATMENT

- Data which is out of normal constraint
- Indicate bad data
- Needs close attention, will have more impact on result
- Can lead to wildly wrong estimations
- Include or exclude based on the impact



Causes of outliers:

- ✓ Artificial
- ✓ Natural

- Data Entry Errors
- Measurement Error
- Experimental Error
- Intentional Outlier
- Data Processing Error
- Sampling error
- Natural Outlier