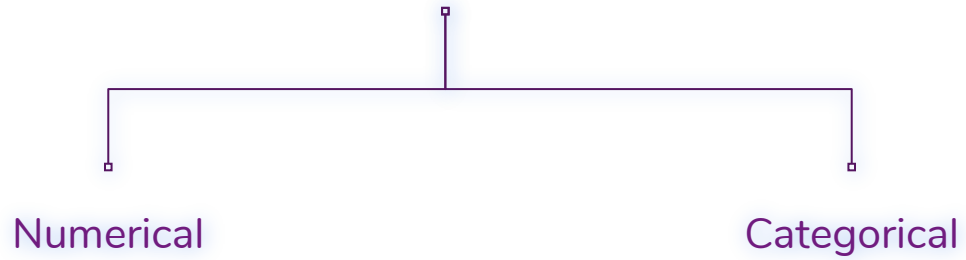
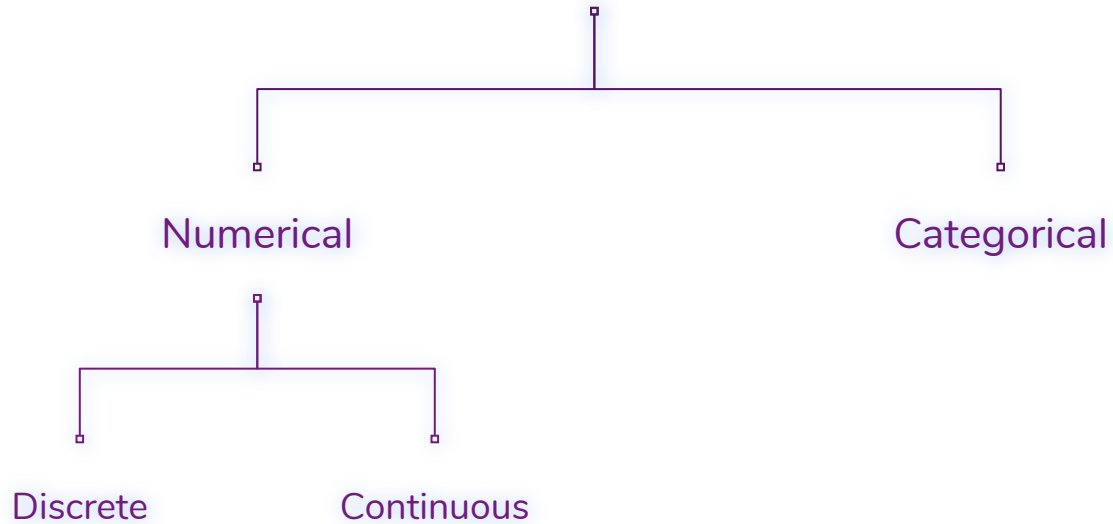


# Summary Statistics

## Types of Data



# Types of Data



# Numerical Data

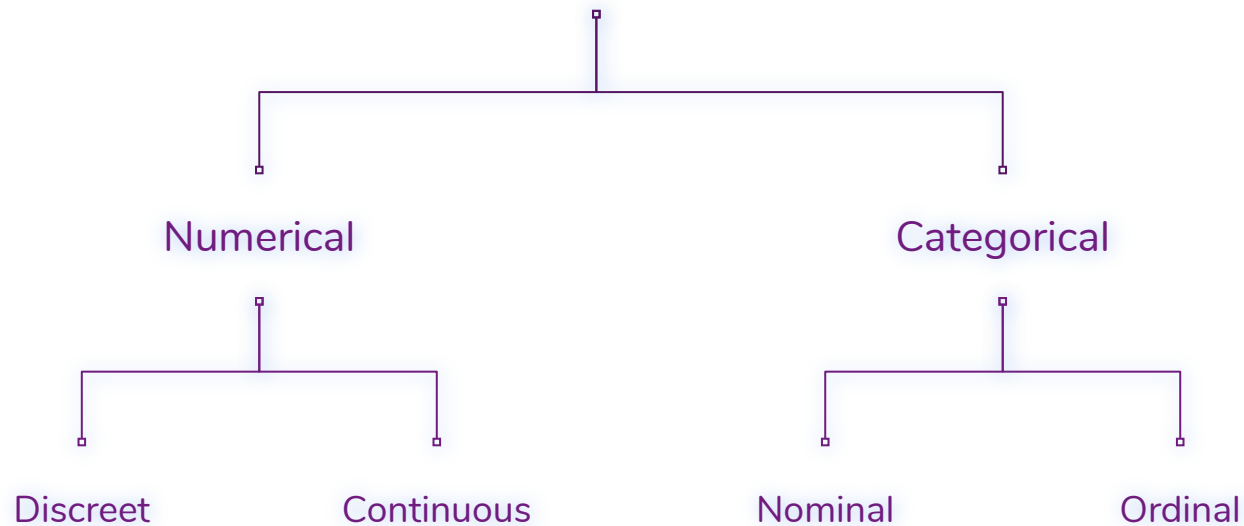
**Discrete** - 1) Takes integer values

2) An example is the number of heads in 100 coin flips

**Continuous** - 1) Takes real values

2) An example would be the height of a person

# Types of Data



# Nominal Data

What is your Gender?

- ☐ Female
- ☐ Male

What languages do you speak?

- ☐ Englisch
- ☐ French
- ☐ German
- ☐ Spanish

# Ordinal Data

What Is Your Educational Background?

- ☐ 1 - Elementary
- ☐ 2 - High School
- ☐ 3 - Undegraduate
- ☐ 4 - Graduate

# Why Data Types

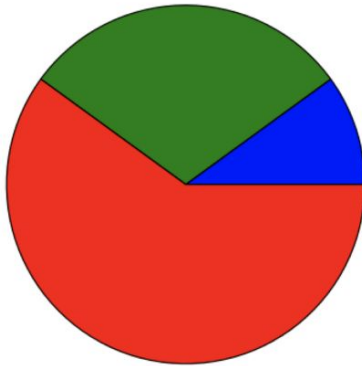
Data Types are an important concept because statistical methods can only be used with certain data types.



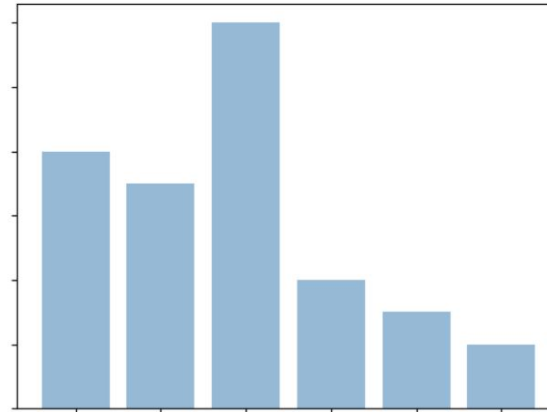
# Nominal Data

Frequencies, Proportion, Percentage. Visualisation Methods: To visualise nominal data you can use a pie chart or a bar chart.

## Pie Chart



## Bar Chart



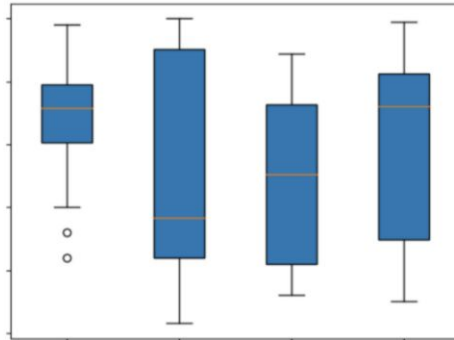
# Ordinal Data

- Frequencies, Proportion, Percentage. Visualisation Methods: To visualise nominal data you can use a pie chart or a bar chart.
- Additionally, you can use percentiles, median, mode and the interquartile range to summarise your data

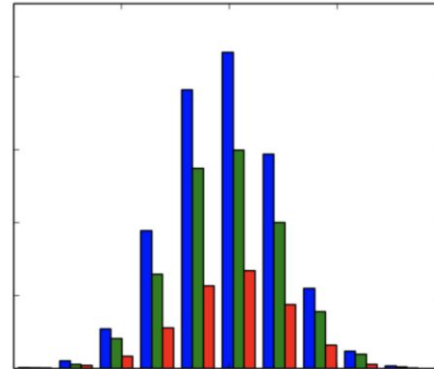
# Continuous Data

Can summarise your data using percentiles, median, interquartile range, mean, mode, standard deviation, and range. Visualis

## Boxplot



## Histogram



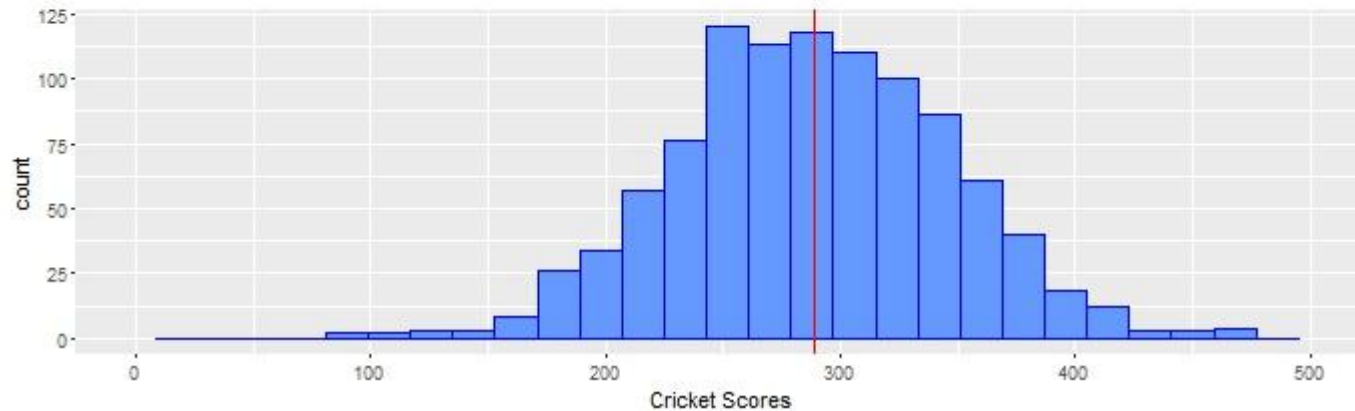
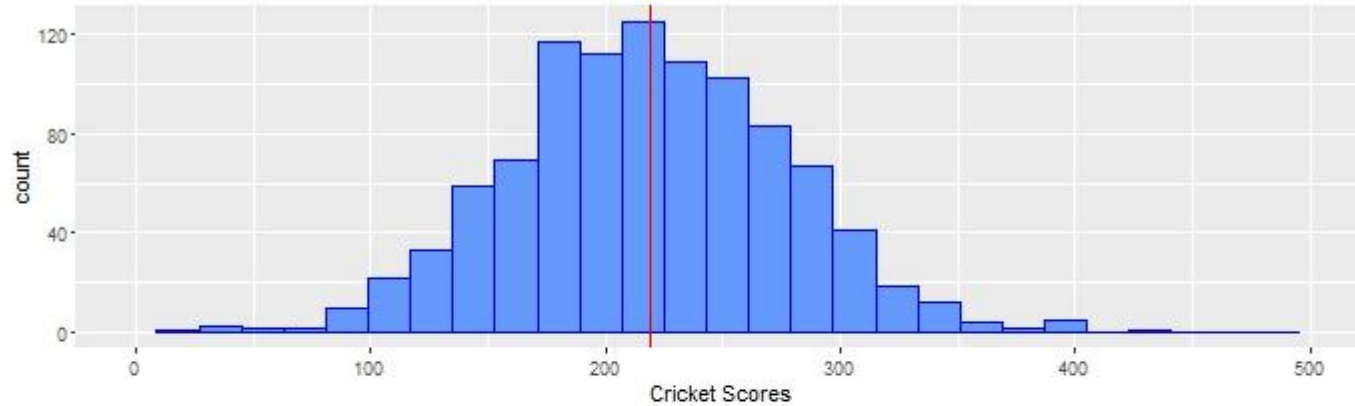
# Summarising data

1. Statistic measures
2. Tables
3. Graphs

# Statistic measures

1. Central Tendency
2. Size / Variability
3. Shape

# Central Tendency - Averages



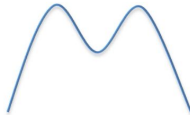
# Central Tendency - Averages

- The mean is simply the average.
- The mode is the value or category that occurs most often within the data
- The median is the “middle” value or midpoint in your data

Unimodal



Bimodal



Multimodal



# Central Tendency - Averages

Q) Find the mean, median and mode of 11, 10, 10, 12, 10, 11, 11 13, 9, 8.



# Central Tendency - Averages

**Q) Find the mean, median and mode of 11, 10, 10, 12, 10, 11, 11 13, 9, 8.**

**Sol:** First sorting the dataset we get 8, 9, 10, 10, 10, 11, 11, 11, 12, 13

**Mean** =  $(8 + 9 + 10 + 10 + 10 + 11 + 11 + 11 + 12 + 13) \div 10 = 105 \div 10 = 10.5$

# Central Tendency - Averages

**Q) Find the mean, median and mode of 11, 10, 10, 12, 10, 11, 11 13, 9, 8.**

**Sol:** First sorting the dataset we get 8, 9, 10, 10, 10, 11, 11, 11, 12, 13

**Mean** =  $(8 + 9 + 10 + 10 + 10 + 11 + 11 + 11 + 12 + 13) \div 10 = 105 \div 10 = 10.5$

**Median** -> middle value ->  $(10 + 1) \div 2 = 5.5$ -th value. The fifth and sixth numbers are the last 10 and the first 11, so:

$$(10 + 11) \div 2 = 21 \div 2 = 10.5$$

# Central Tendency - Averages

**Q) Find the mean, median and mode of 11, 10, 10, 12, 10, 11, 11 13, 9, 8.**

**Sol:** First sorting the dataset we get 8, 9, 10, 10, 10, 11, 11, 11, 12, 13

**Mean** =  $(8 + 9 + 10 + 10 + 10 + 11 + 11 + 11 + 12 + 13) \div 10 = 105 \div 10 = 10.5$

**Median** -> middle value ->  $(10 + 1) \div 2 = 5.5$ -th value. The fifth and sixth numbers are the last 10 and the first 11, so:

$$(10 + 11) \div 2 = 21 \div 2 = 10.5$$

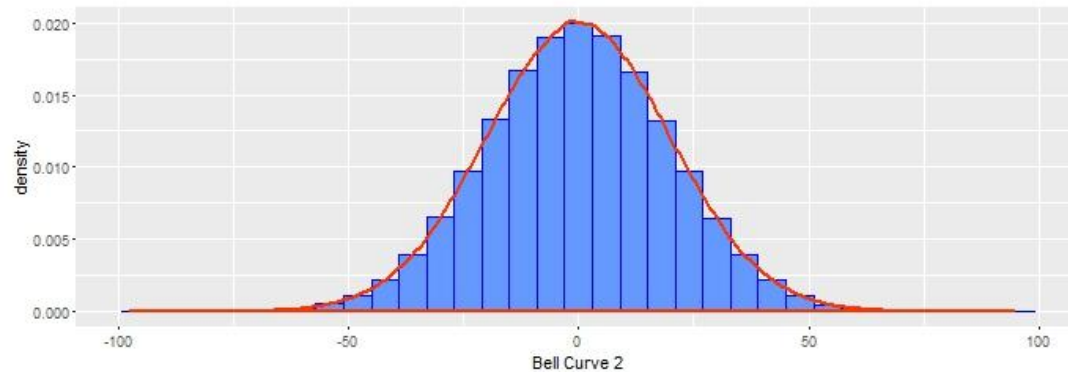
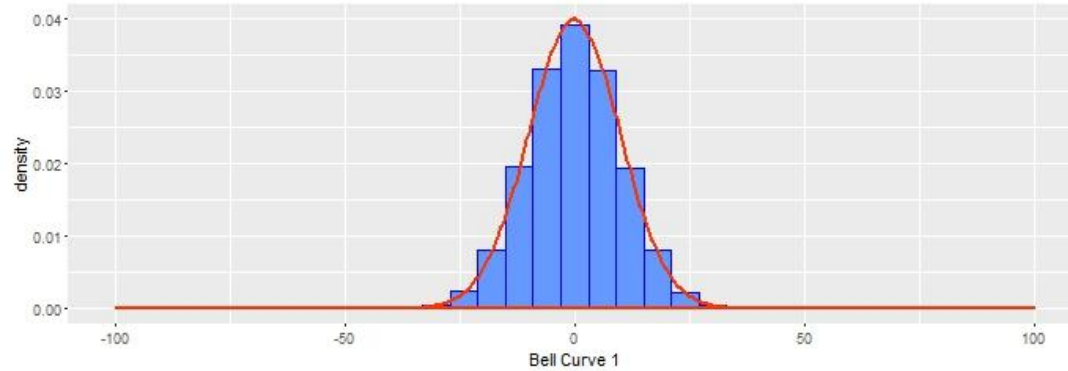
**Mode** - This list has two values that are repeated three times; namely, 10 and 11,

Mean: 10.5, Median: 10.5, Modes: 10 and 11

# Spread/Variability

- Variance
- Standard deviation
- Range
- Interquartile range (IQR)

# Spread/Variability



# Spread/Variability

- Variance
- Standard deviation
- Range
- Inter quartile range (IQR)

# Variance & Standard deviation

- Variance ( $S^2$ ) = average squared deviation of values from mean
- Standard deviation ( $S$ ) = square root of the variance

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

where  $S$  = the standard deviation of a sample,  
 $\Sigma$  means "sum of,"  
 $X$  = each value in the data set,  
 $\bar{X}$  = mean of all values in the data set,  
 $N$  = number of values in the data set.

# Variance & Standard deviation

Q) You have four dogs with heights of 600mm, 470mm, 170mm, 430mm and 300mm. What's the variance and standard deviation?



# Variance & Standard deviation

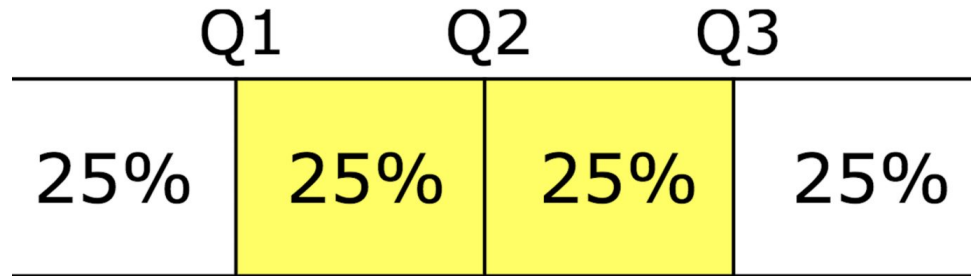
Sol : First calculate mean  $(600 + 470 + 170 + 430 + 300)/5 = 394$

$$\begin{aligned}\text{Variance} &= \frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5} \\ &= \frac{42436 + 5776 + 50176 + 1296 + 8836}{5} \\ &= \frac{108520}{5} \\ &= 21704\end{aligned}$$

$$\begin{aligned}\text{Standard deviation} &= \sqrt{21704} \\ &= 147.32...\end{aligned}$$

# Range & IQR

- Range
- Interquartile range (IQR)



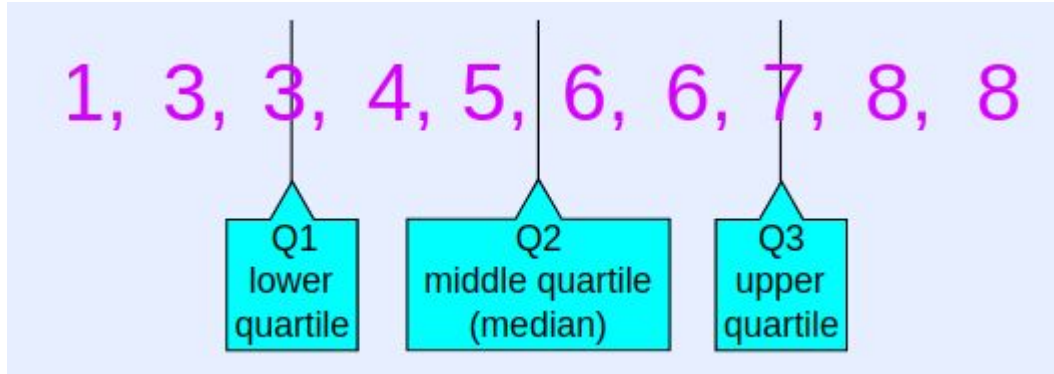
Interquartile Range  
=  $Q3 - Q1$

# Range & IQR

Q) Find range & interquartile range for 1, 3, 3, 4, 5, 6, 6, 7, 8, 8

# Range & IQR

Sol:



Quartile 1 (Q1) = 3

Quartile 2 (Q2) = 5.5

Quartile 3 (Q3) = 7

Range = Max - Min = 8 - 1 = 7

IQR = Q3 - Q1 = 7 - 3 = 4

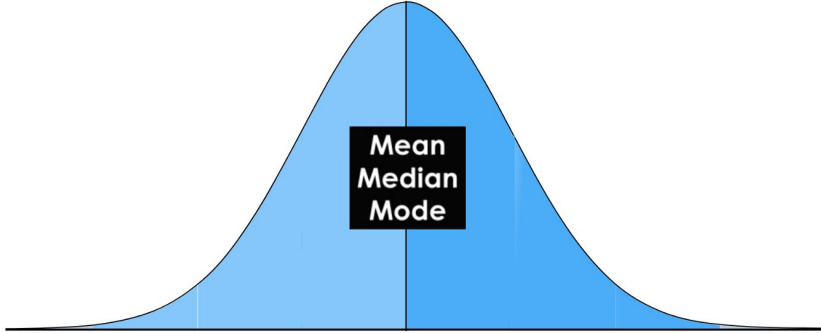
# Spread/Variability

- Variance
- Standard deviation
- Range
- Interquartile range (IQR)

# Shape

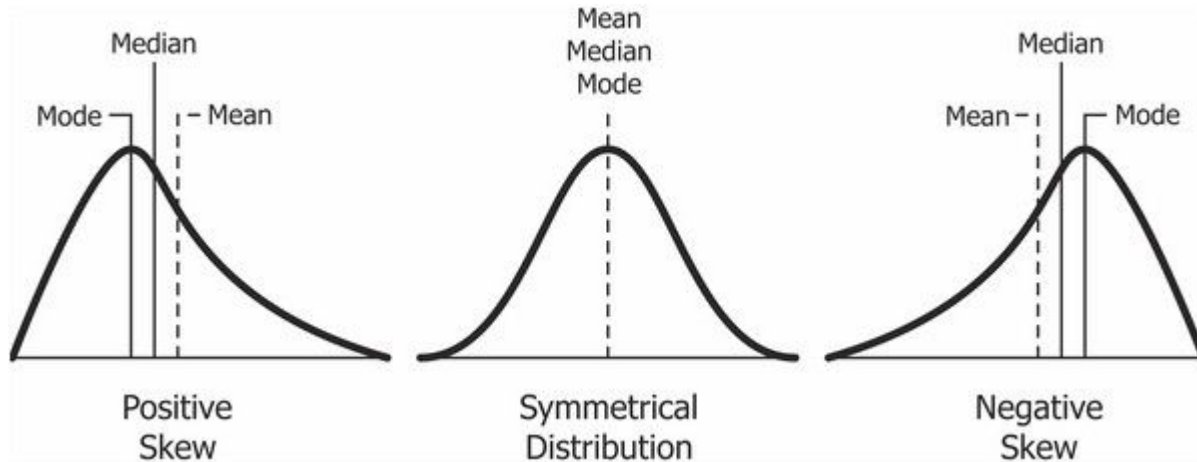
- Skewness
- Kurtosis

# Standard Bell Curve



# Skewness

- Skewness



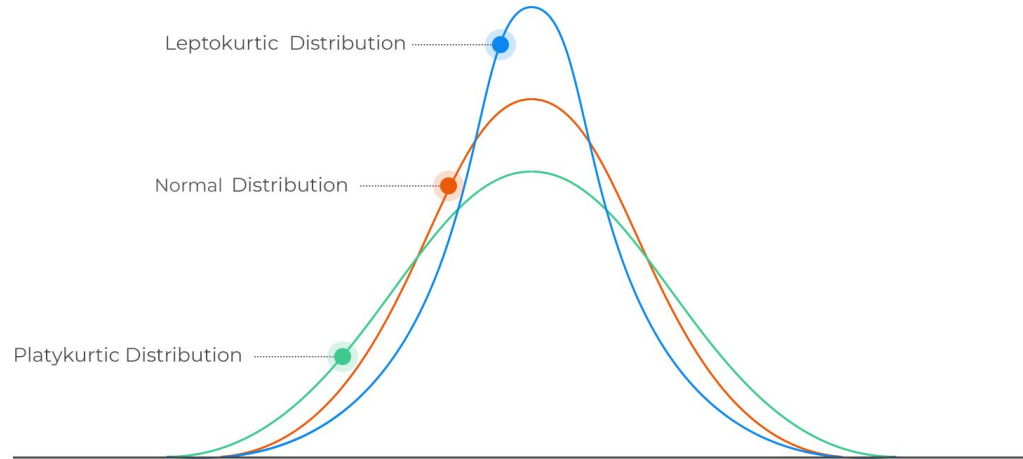


# Kurtosis

Kurtosis measures whether your dataset is heavy-tailed or light-tailed compared to a normal distribution



## Kurtosis



# Summarising data

1. Statistic measures
2. Tables
3. Graphs

# Tables

Class Interval	Class Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
60–64	1	$\frac{1}{25} = 0.04$	1	0.04
65–69	1	$\frac{1}{25} = 0.04$	$1 + 1 = 2$	$0.04 + 0.04 = 0.08$
70–74	2	$\frac{2}{25} = 0.08$	$2 + 2 = 4$	$0.08 + 0.08 = 0.16$
75–79	6	$\frac{6}{25} = 0.24$	$4 + 6 = 10$	$0.16 + 0.24 = 0.4$
80–84	3	$\frac{3}{25} = 0.12$	$10 + 3 = 13$	$0.4 + 0.12 = 0.52$
85–89	5	$\frac{5}{25} = 0.2$	$13 + 5 = 18$	$0.52 + 0.2 = 0.72$
90–94	5	$\frac{5}{25} = 0.2$	$18 + 5 = 23$	$0.72 + 0.2 = 0.92$
95–99	2	$\frac{2}{25} = 0.08$	$23 + 2 = 25$	$0.92 + 0.08 = 1$

# Summarising data

1. Statistic measures
2. Tables
3. Graphs

# Graphs

- Pie
- Bar
- Histogram
- Line

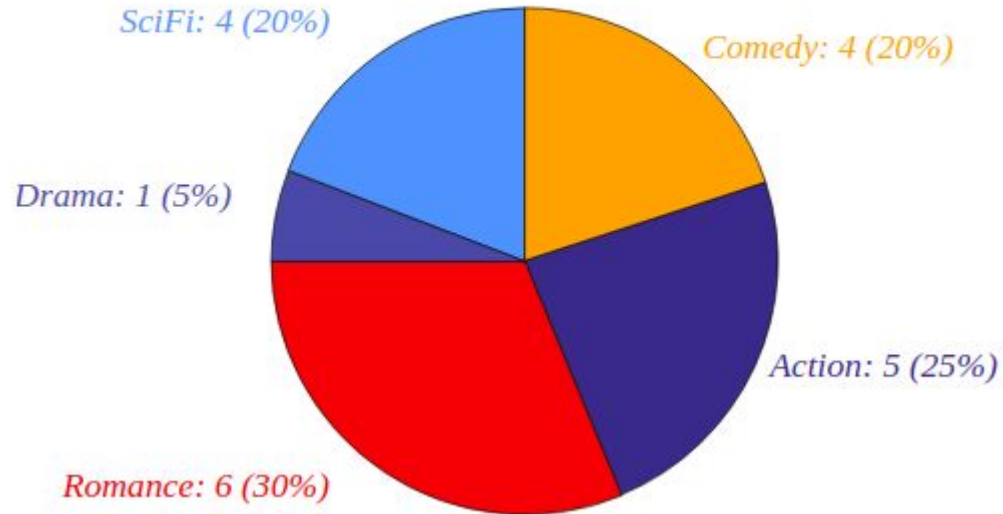
# Pie

*Table: Favorite Type of Movie*

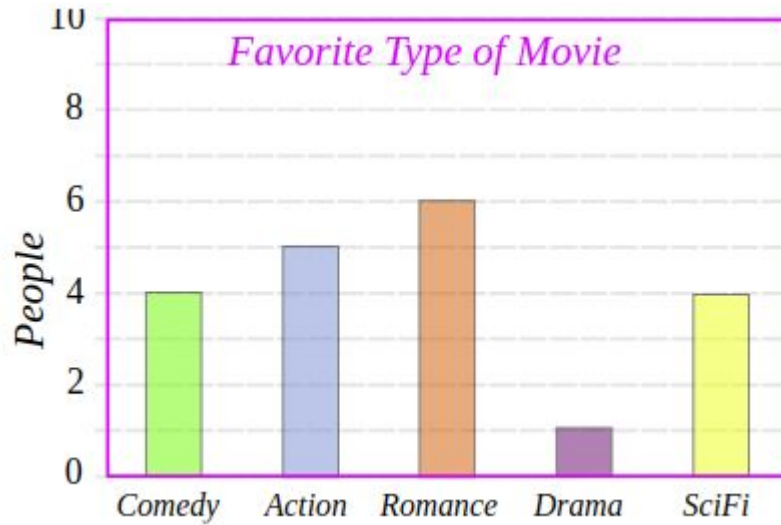
Comedy	Action	Romance	Drama	SciFi
4	5	6	1	4

# Pie

*Favorite Type of Movie*

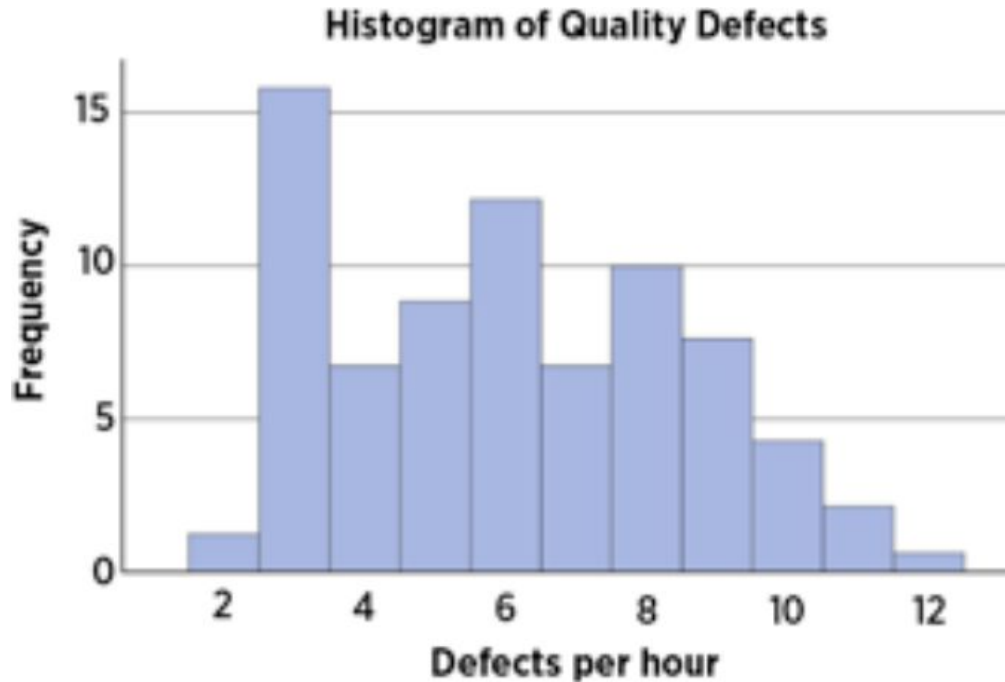


# Bar





# Histogram



# Line graph

Produce sales  
IN THOUSANDS (USD)



# Summary

1. Statistic measures
2. Tables
3. Graphs

# Summary

1. Statistic measures - Central tendency, Spread, Shape
2. Tables
3. Graphs

# Summary

1. Statistic measures - **Central tendency**(mean, median, mode), Spread, Shape
2. Tables
3. Graphs

# Summary

1. Statistic measures - Central tendency(mean, median, mode), Spread(variance, standard deviation, range, IQR), Shape
2. Tables
3. Graphs

# Summary

1. Statistic measures - Central tendency(mean, median, mode), Spread(variance, standard deviation, range, IQR), **Shape**(skewness, kurtosis)
2. Tables
3. Graphs

# Summary

1. Statistic measures - Central tendency(mean, median, mode), Spread(variance, standard deviation, range, IQR), Shape(skewness, kurtosis)
2. **Tables** - Frequency, Cumulative Frequency
3. Graphs



# Summary

1. Statistic measures - Central tendency(mean, median, mode), Spread(variance, standard deviation, range, IQR), Shape(skewness, kurtosis)
2. Tables - Frequency, Cumulative Frequency
3. **Graphs** - Pie, Bar, Histogram, Line