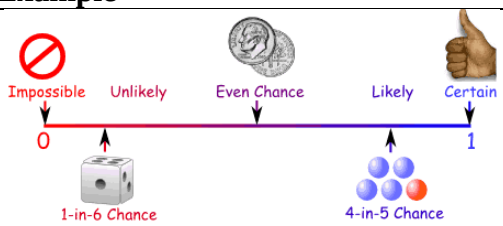
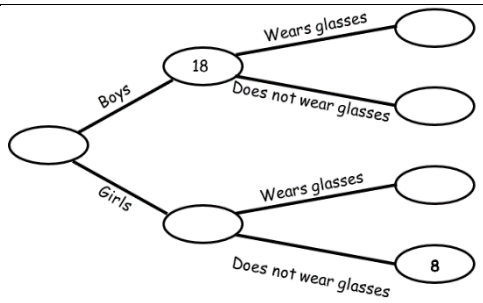
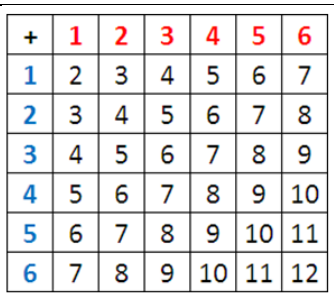
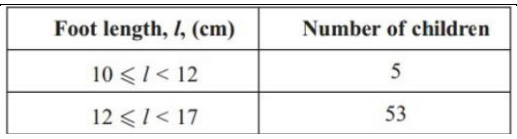
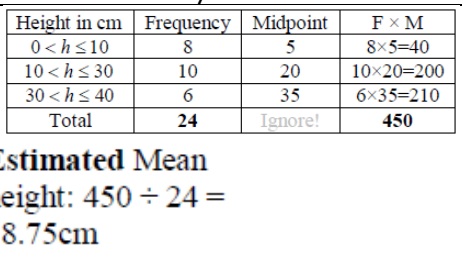
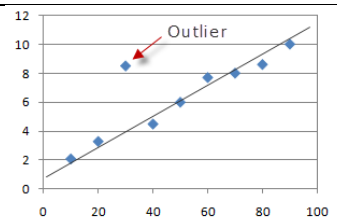
Topic: Basic Probability

Topic/Skill	Definition/Tips	Example
1. Probability	The **likelihood/chance** of something happening. Is expressed as a number **between 0 (impossible) and 1 (certain)**. Can be expressed as a fraction, decimal, percentage or in words (likely, unlikely, even chance etc.)	
2. Probability Notation	**P(A)** refers to the **probability that event A will occur**.	P(Red Queen) refers to the probability of picking a Red Queen from a pack of cards.
3. Theoretical Probability	$\frac{\textit{Number of Favourable Outcomes}}{\textit{Total Number of Possible Outcomes}}$	Probability of rolling a 4 on a fair 6-sided die = $\frac{1}{6}$.
4. Relative Frequency	$\frac{\textit{Number of Successful Trials}}{\textit{Total Number of Trials}}$	A coin is flipped 50 times and lands on Tails 29 times. The relative frequency of getting Tails = $\frac{29}{50}$.
5. Expected Outcomes	To find the number of expected outcomes, **multiply** the **probability** by the **number of trials**.	The probability that a football team wins is 0.2 How many games would you expect them to win out of 40? $0.2 \times 40 = 8\ games$
6. Exhaustive	Outcomes are **exhaustive** if they **cover the entire range of possible outcomes**. The **probabilities** of an **exhaustive** set of outcomes **adds up to 1**.	When rolling a six-sided die, the outcomes 1, 2, 3, 4, 5 and 6 are exhaustive, because they cover all the possible outcomes.
7. Mutually Exclusive	Events are mutually exclusive if they **cannot happen at the same time**. The **probabilities** of an exhaustive set of **mutually exclusive** events **adds up to 1**.	Examples of mutually exclusive events: - Turning left and right - Heads and Tails on a coin Examples of non mutually exclusive events: - King and Hearts from a deck of cards, because you can pick the King of Hearts
8. Frequency Tree	A diagram showing how information is categorised into various categories. The **numbers** at the ends of branches tells us how often something happened (**frequency**).	

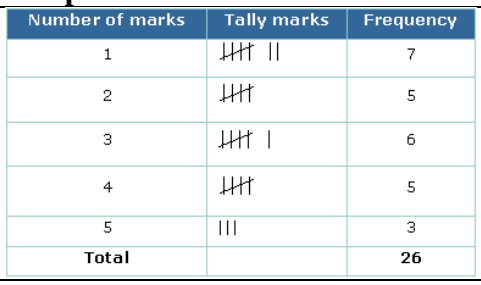
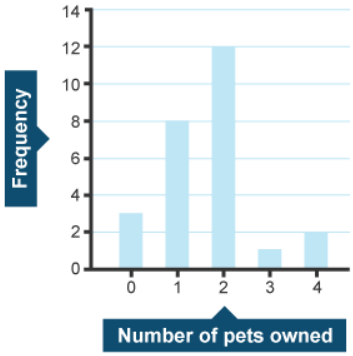
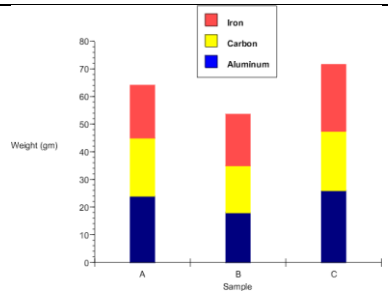
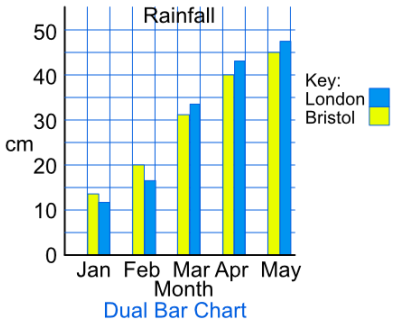
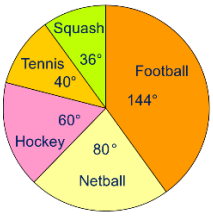
	The **lines** connected the numbers are called **branches**.	
9. Sample Space	The **set of all possible outcomes** of an experiment.	
10. Sample	A **sample** is a small selection of items from a population. A sample is **biased** if individuals or groups from the population are not represented in the sample.	A sample could be selecting 10 students from a year group at school.
11. Sample Size	The larger a sample size, the closer those probabilities will be to the true probability.	A sample size of 100 gives a more reliable result than a sample size of 10.

Topic: Summarising Data

Topic/Skill	Definition/Tips	Example
1. Types of Data	**Qualitative** Data – **non-numerical** data **Quantitative** Data – **numerical** data **Continuous** Data – data that can take **any numerical value** within a given range. **Discrete** Data – data that can take **only specific values** within a given range.	Qualitative Data – eye colour, gender etc. Continuous Data – weight, voltage etc. Discrete Data – number of children, shoe size etc.
2. Grouped Data	Data that has been **bundled in to categories**. Seen in grouped frequency tables, histograms, cumulative frequency etc.	 Foot length, l, (cm) | Number of children $10 \leqslant l < 12$ | 5 $12 \leqslant l < 17$ | 53
3. Primary /Secondary Data	**Primary** Data – **collected yourself** for a specific purpose. **Secondary** Data – **collected by someone else** for another purpose.	Primary Data – data collected by a student for their own research project. Secondary Data – Census data used to analyse link between education and earnings.
4. Mean	**Add** up the values and **divide** by how many values there are.	The mean of 3, 4, 7, 6, 0, 4, 6 is $\frac{3+4+7+6+0+4+6}{7} = 5$
5. Mean from a Table	1. Find the midpoints (if necessary) 2. Multiply Frequency by values or midpoints 3. Add up these values 4. Divide this total by the Total Frequency If **grouped** data is used, the answer will be an **estimate**.	 Height in cm | Frequency | Midpoint | F × M $0 < h \leq 10$ | 8 | 5 | 8×5=40 $10 < h \leq 30$ | 10 | 20 | 10×20=200 $30 < h \leq 40$ | 6 | 35 | 6×35=210 Total | **24** | Ignore! | **450** **Estimated** Mean height: 450 ÷ 24 = 18.75cm
6. Median Value	The **middle** value. Put the data in order and find the middle one. If there are **two middle values**, find the number half way between them by **adding them together and dividing by 2**.	Find the median of: 4, 5, 2, 3, 6, 7, 6 Ordered: 2, 3, 4, **5**, 6, 6, 7 Median = 5
7. Median from a Table	Use the formula $\frac{(n+1)}{2}$ to find the position of the median. n is the total frequency.	If the total frequency is 15, the median will be the $\left(\frac{15+1}{2}\right) = 8th$ position
8. Mode /Modal Value	**Most** frequent/common. Can have more than one mode (called bi-modal or multi-modal) or no mode (if all values appear once)	Find the mode: 4, 5, 2, 3, 6, 4, 7, 8, 4 Mode = 4
9. Range	**Highest value subtract the Smallest value**	Find the range: 3, 31, 26, 102, 37, 97. Range = 102-3 = 99

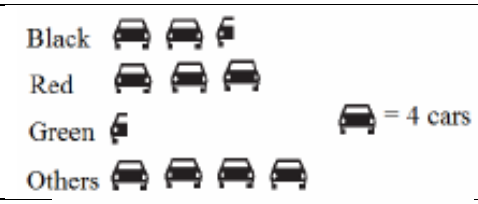
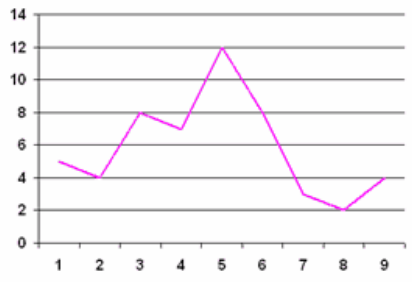
	Range is a 'measure of spread'. The smaller the range the more <u>consistent</u> the data.	
10. Outlier	A value that '**lies outside**' most of the other values in a set of data. An outlier is **much smaller or much larger** than the other values in a set of data.	
11. Lower Quartile	**Divides** the **bottom half** of the data into **two halves**. **LQ = $Q_1 = \frac{(n+1)}{4}$ th value**	Find the lower quartile of: 2, **<u>3</u>**, 4, 5, 6, 6, 7 $Q_1 = \frac{(7+1)}{4} = 2nd$ value → 3
12. Lower Quartile	**Divides** the **top half** of the data into **two halves**. **UQ = $Q_3 = \frac{3(n+1)}{4}$ th value**	Find the upper quartile of: 2, 3, 4, 5, 6, **<u>6</u>**, 7 $Q_3 = \frac{3(7+1)}{4} = 6th$ value → 6
13. Interquartile Range	The **difference** between the **upper quartile and lower quartile.** $$IQR = Q_3 - Q_1$$ The **smaller** the **interquartile range**, the **more consistent** the data.	Find the IQR of: 2, 3, 4, 5, 6, 6, 7 $$IQR = Q_3 - Q_1 = 6 - 3 = 3$$

Topic: Representing Data

Topic/Skill	Definition/Tips	Example
1. Frequency Table	A record of **how often each value** in a set of data **occurs**.	
2. Bar Chart	Represents data as vertical blocks. $\boldsymbol{x-axis}$ shows the **type** of data $\boldsymbol{y-axis}$ shows the **frequency** for each type of data Each bar should be the **same width** There should be **gaps** between each bar Remember to **label** each axis.	
3. Types of Bar Chart	**Compound/Composite** Bar Charts show data stacked on top of each other. **Comparative/Dual** Bar Charts show data side by side.	 
4. Pie Chart	Used for showing **how data breaks down into** its constituent **parts**. When drawing a pie chart, **divide 360° by the total frequency**. This will tell you how many degrees to use for the frequency of each category. Remember to **label** the category that each sector in the pie chart represents.	 If there are 40 people in a survey, then each person will be worth 360÷40=9° of the pie chart.

Number of marks	Tally marks	Frequency
1	𝍸 ||	7
2	𝍸	5
3	𝍸 |	6
4	𝍸	5
5	|||	3
Total		26

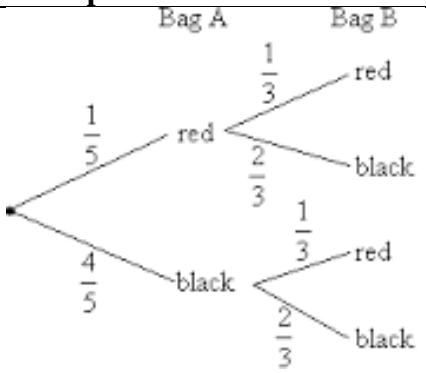
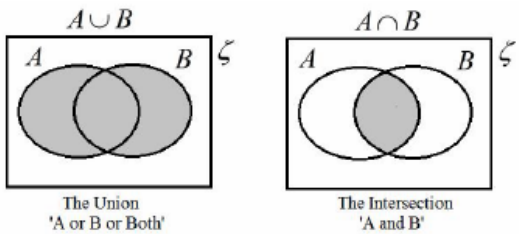
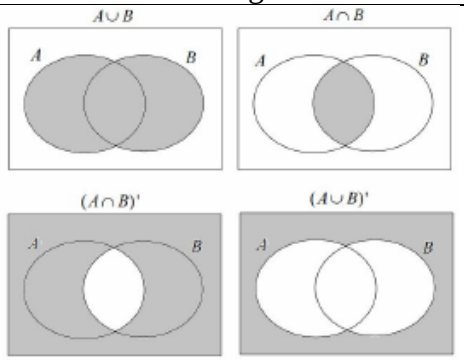
Dual Bar Chart

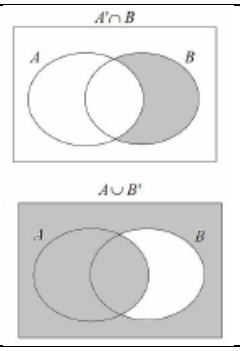
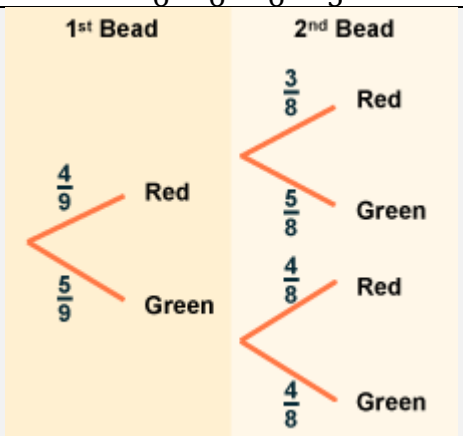
5. Pictogram	Uses **pictures** or symbols to **show the value** of the data. A pictogram must have a **key**.	
6. Line Graph	A graph that uses **points connected by straight lines** to show how data changes in values. This can be used for **time series data**, which is a series of data points spaced over uniform time intervals in **time order**.	
7. Two Way Tables	A table that **organises data** around **two categories.** Fill out the information step by step using the information given. Make sure all the totals add up for all columns and rows.	**Question: Complete the 2 way table below.** Left Handed / Right Handed / Total Boys: 10 / / 58 Girls: / / / Total: / 84 / 100 **Answer: Step 1, fill out the easy parts (the totals)** Boys: 10 / *48* / 58 Girls: / / *42* Total: *16* / 84 / 100 **Answer: Step 2, fill out the remaining parts** Boys: 10 / *48* / 58 Girls: *6* / *36* / *42* Total: *16* / 84 / 100
8. Box Plots	The minimum, lower quartile, median, upper quartile and maximum are shown on a box plot. A box plot can be drawn independently or from a cumulative frequency diagram.	Students sit a maths test. The highest score is 19, the lowest score is 8, the median is 14, the lower quartile is 10 and the upper quartile is 17. Draw a box plot to represent this information. 
9. Comparing Box Plots	Write two sentences. 1. Compare the **averages** using the **medians** for two sets of data. 2. Compare the **spread** of the data using the **range or IQR** for two sets of data. The <u>smaller</u> the range/IQR, the <u>more consistent</u> the data. You must compare box plots **in the context of the problem**.	'On average, students in class A were more successful on the test than class B because their median score was higher.' 'Students in class B were more consistent than class A in their test scores as their IQR was smaller.'

Topic: Systematic Listing

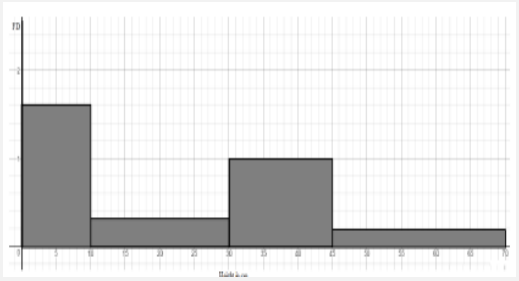
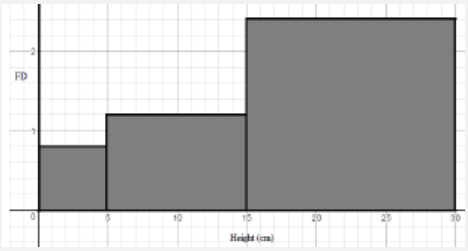
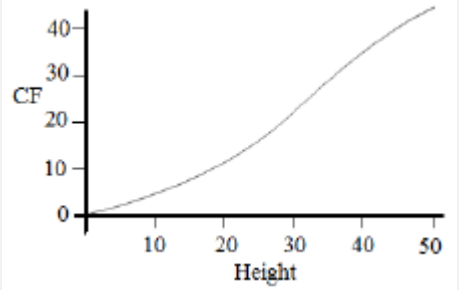
Topic/Skill	Definition/Tips	Example
1. Combination	A collection of things, where the **order does not matter**.	How many combinations of two ingredients can you make with apple, banana and cherry? Apple, Banana Apple, Cherry Banana, Cherry 3 combinations
2. Permutation	A collection of things, where the **order does matter**.	You want to visit the homes of three friends, Alex (A), Betty (B) and Chandra (C) but haven't decided the order. What choices do you have? ABC ACB BAC BCA CAB CBA
3. Permutations with Repetition	When something has n different types, there are $\boldsymbol{n}$ **choices each time**. Choosing r of something that has n different types, the permutations are: $n \times n \times \ldots (r\ times) = \boldsymbol{n^r}$	How many permutations are there for a three-number combination lock? 10 numbers to choose from $\{1, 2, \ldots .10\}$ and we choose 3 of them → $10 \times 10 \times 10 = 10^3 = 1000$ permutations.
4. Permutations without Repetition	We have to **reduce the number of available choices each time**. One you have chosen something, you cannot choose it again.	How many ways can you order 4 numbered balls? $4 \times 3 \times 2 \times 1 = 24$
5. Factorial	The factorial symbol '!' means to multiply a series of descending integers to 1. Note: $0! = 1$	$4! = 4 \times 3 \times 2 \times 1 = 24$
6. Product Rule for Counting	If there are $\boldsymbol{x}$ **ways of doing something** and $\boldsymbol{y}$ **ways of doing something else**, then there are $\boldsymbol{xy}$ **ways of performing both**.	To choose one of $\{A, B, C\}$ and one of $\{X, Y\}$ means to choose one of $\{AX, AY, BX, BY, CX, CY\}$ The rule says that there are $3 \times 2 = 6$ choices.

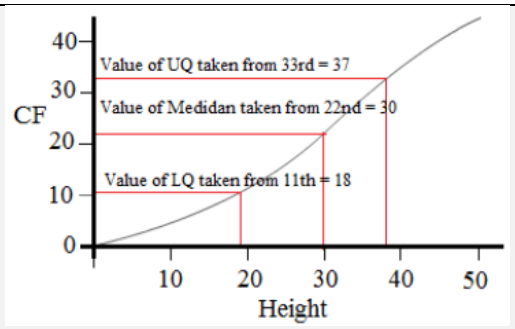
Topic: Probability (Trees and Venns)

Topic/Skill	Definition/Tips	Example
1. Tree Diagrams	Tree diagrams show **all the possible outcomes** of an event and calculate their probabilities. **All branches must add up to 1 when adding downwards.** This is because the **probability of something not happening** is **1 minus the probability that it does happen**. **Multiply** going **across** a tree diagram. **Add** going **down** a tree diagram.	
2. Independent Events	The outcome of a **previous event does not influence/affect the outcome of a second event**.	An example of independent events could be <u>replacing</u> a counter in a bag after picking it.
3. Dependent Events	The outcome of a **previous event does influence/affect the outcome of a second event**.	An example of dependent events could be not replacing a counter in a bag after picking it. '<u>Without replacement</u>'
4. Probability Notation	**P(A)** refers to the **probability that event A will occur**. **P(A')** refers to the **probability that event A will <u>not</u> occur**. **P(A ∪ B)** refers to the **probability that event A <u>or</u> B <u>or</u> both will occur.** **P(A ∩ B)** refers to the **probability that <u>both</u> events A and B will occur.**	P(Red Queen) refers to the probability of picking a Red Queen from a pack of cards. P(Blue') refers to the probability that you do not pick Blue. P(Blonde ∪ Right Handed) refers to the probability that you pick someone who is Blonde or Right Handed or both. P(Blonde ∩ Right Handed) refers to the probability that you pick someone who is both Blonde and Right Handed.
5. Venn Diagrams	A Venn Diagram shows the **relationship between a group of different things** and how they overlap. You may be asked to shade Venn Diagrams as shown below and to the right. 	

		
6. Venn Diagram Notation	∈ means '**element of a set**' (a value in the set) { } means the collection of values in the set. ξ means the '**universal set**' (all the values to consider in the question) **A' means 'not in set A' (called complement)** **A ∪ B means 'A or B or both' (called Union)** **A ∩ B means 'A and B (called Intersection)**	Set A is the even numbers less than 10. A = {2, 4, 6, 8} Set B is the prime numbers less than 10. B = {2, 3, 5, 7} A ∪ B = {2, 3, 4, 5, 6, 7, 8} A ∩ B = {2}
7. AND rule for Probability	When two events, A and B, are **independent**: $\boldsymbol{P(A \text{ and } B) = P(A) \times P(B)}$	What is the probability of rolling a 4 and flipping a Tails? $P(4 \text{ and } Tails) = P(4) \times P(Tails)$ $= \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$
8. OR rule for Probability	When two events, A and B, are **mutually exclusive**: $\boldsymbol{P(A \text{ or } B) = P(A) + P(B)}$	What is the probability of rolling a 2 or rolling a 5? $P(2 \text{ or } 5) = P(2) + P(5)$ $= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$
9. Conditional Probability	The probability of an event A happening, **given that** event B has already happened. With conditional probability, check if the numbers on the second branches of a tree diagram changes. For example, if you have 4 red beads in a bag of 9 beads and pick a red bead on the first pick, then there will be 3 red beads left out of 8 beads on the second pick.	

Topic: Histograms and Cumulative Frequency

Topic/Skill	Definition/Tips	Example
1. Histograms	A visual way to display frequency data using bars. Bars can be **unequal in width**. Histograms show **frequency density** on the **y-axis**, not frequency. $$\textbf{Frequency Density} = \frac{\textbf{Frequency}}{\textbf{Class Width}}$$ | Height(cm) | Frequency | | $0 < h \leq 10$ | 8 | | $10 < h \leq 30$ | 6 | | $30 < h \leq 45$ | 15 | | $45 < h \leq 70$ | 5 |	| Frequency Density (FD) | | $8 \div 5 = 1.6$ | | $6 \div 20 = 0.3$ | | $15 \div 15 = 1$ | | $5 \div 25 = 0.2$ | 
2. Interpreting Histograms	The **area** of the bar is proportional to the **frequency** of that class interval. $$\textbf{Frequency} = \textbf{Freq Density} \times \textbf{Class Width}$$	A histogram shows information about the heights of a number of plants. 4 plants were less than 5cm tall. Find the number of plants more than 5cm tall.  Above 5cm: 1.2 x 10 + 2.4 x 15 = 12 + 36 = 48
3. Cumulative Frequency	Cumulative Frequency is a **running total**. | Age | Frequency | | $0 < a \leq 10$ | 15 | | $10 < a \leq 40$ | 35 | | $40 < a \leq 50$ | 10 |	| Cumulative Frequency | | 15 | | $15 + 35 = 50$ | | $50 + 10 = 60$ |
4. Cumulative Frequency Diagram	A cumulative frequency diagram is a **curve that goes up**. It looks a little like a stretched-out **S shape**. Plot the cumulative frequencies at the **end-point** of each interval.	

5. Quartiles from Cumulative Frequency Diagram	**Lower Quartile** (Q1): **25%** of the data is less than the lower quartile. **Median** (Q2): **50%** of the data is less than the median. **Upper Quartile** (Q3): **75%** of the data is less than the upper quartile. **Interquartile Range (IQR)**: represents the **middle 50%** of the data.	 $IQR = 37 - 18 = 19$
6. Hypothesis	**A statement that might be true, which can be tested.**	Hypothesis: 'Large dogs are better at catching tennis balls than small dogs'. We can test this hypothesis by having hundreds of different sized dogs try to catch tennis balls.