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.	<table><tr><th>Foot length, l, (cm)</th><th>Number of children</th></tr><tr><td>$10 \leqslant l < 12$</td><td>5</td></tr><tr><td>$12 \leqslant l < 17$</td><td>53</td></tr></table>
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**.	<table><tr><th>Height in cm</th><th>Frequency</th><th>Midpoint</th><th>F × M</th></tr><tr><td>$0 < h \leq 10$</td><td>8</td><td>5</td><td>8×5=40</td></tr><tr><td>$10 < h \leq 30$</td><td>10</td><td>20</td><td>10×20=200</td></tr><tr><td>$30 < h \leq 40$</td><td>6</td><td>35</td><td>6×35=210</td></tr><tr><td>Total</td><td>24</td><td>Ignore!</td><td>450</td></tr></table> **Estimated** Mean height: $450 \div 24 = 18.75$cm
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**. $$LQ = Q_1 = \frac{(n+1)}{4} \, th \text{ value}$$	Find the lower quartile of: 2, **3**, 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**. $$UQ = Q_3 = \frac{3(n+1)}{4} \, th \text{ value}$$	Find the upper quartile of: 2, 3, 4, 5, 6, **6**, 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 $$IQR = Q_3 - Q_1 = 6 - 3 = 3$$