Math Hawker Digest **Secondary 4**

Elementary Mathematics – Visualization

1. Categorical Data **2. Discrete Numbers** **3. Continuous Numbers**

1 Categorical Data

We often group information into categories to make sense of the world. Visualising these categories helps us draw deeper insights.

Frequency Table

A simple table can organize categorical* data. Listing the frequency of each value shows how often it occurs and helps reveal patterns in the data.

Discrete Values	A	B	C	D	E
Counts	34	90	56	12	78

Pros

- Simple tool for organising raw data.
- Easy to construct and interpret, difficult to misuse.

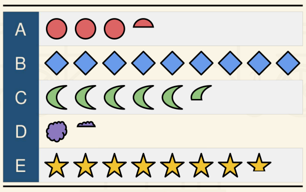
Cons

- Does not clearly show the data distribution and trends.
- Less visual and engaging.
- Can become lengthy or hard to analyze with large datasets.

*we'll summarize the different types of data in the glossary at the end of these notes.

Pictogram

A pictogram uses icons or pictures to represent counts. Partial icons can be used to show non-whole values.



Pros (vs Frequency Table)

- More visual, engaging, and memorable.
- Icons make differences in categories more obvious at a glance.

Cons

- Partial icons are difficult to precisely interprate.
- Can be misleading if icons are of different sizes.

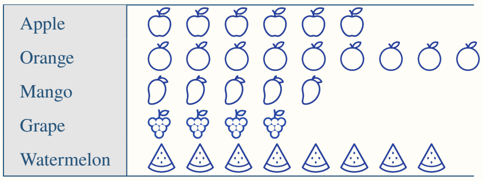
Example 1:

A juice stall recorded the number of cups sold for each flavor over one week. The results are shown in the table below.

Flavor	Apple	Orange	Mango	Grape	Watermelon
Sold	30	45	25	20	40

a) Create a pictogram based on the data in the table.
b) Which flavor was the most popular?
c) What percentage of the total cups sold were Mango juice?

a)



Each symbol represents 5 cups sold.

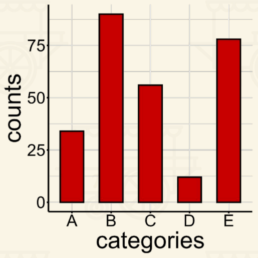
b) – The most popular flavor is Orange.

c) – $\frac{5}{6+9+5+4+8} \times 100\% = 15.6\%$

– 15.6% of the total cups sold were Mango juice.

Bar Graph

A bar graph shows categorical data using bars of equal width. The height of each bar represents continuous data (often counts). The x-axis generally do not need to be in any specific order.



Pros (vs Pictogram)

- More precise and easily interpreted.
- Easier to construct and takes up less space.
- Works with both counts (number of fruits sold) and relative data (temperature per month).

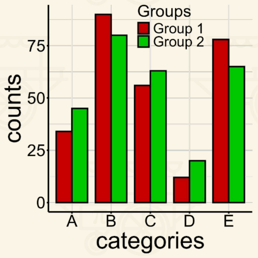
Cons

- Less visually engaging and memorable.
- Axis truncation* can lead to misinterpretation.

*axis truncation happens when a graph doesn't start the vertical (y-) axis at zero. This can make small differences look much bigger than they really are. This can be a strength depending on intended message but opens up possibility for misuse.

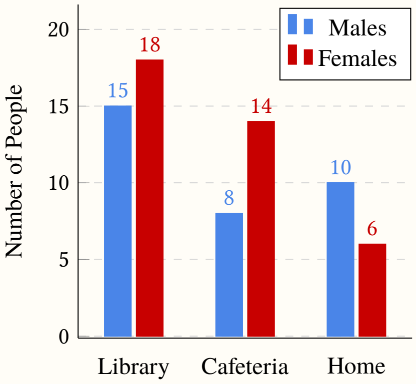
Grouped Bar Graph

Grouped bar graphs enable the easy comparison of multiple datasets that share the same categories.



Example 1:

A school surveyed students on their preferred study locations. The results were stratified by gender.



a) How many students preferred to study in the Library?
b) Which location had the greatest gender difference in preference?
c) What percentage of surveyed students preferred to study at Home?
d) Describe one trend you notice in how study preferences differ between male and female students.

a) – $12 + 18 = 30$
– 30 students preferred to study in the Library.

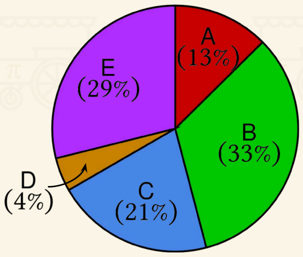
b) – $14 - 8 = 6 > 10 - 6 = 4 > 18 - 15 = 3$
– The Cafeteria had the largest gender difference in preference.

c) – $\frac{10 + 6}{15 + 18 + 8 + 14 + 10 + 6} \times 100\% = 22.5\%$
– 22.5% of all students preferred to study at Home.

d) – Male students prefer to study at Home than outside.

Pie Chart

A pie chart shows proportions of a whole using slices of a circle. It's ideal for displaying how a dataset is divided into non-overlapping categories, and all parts should add up to 100%.



Pros (vs Bar Graph)

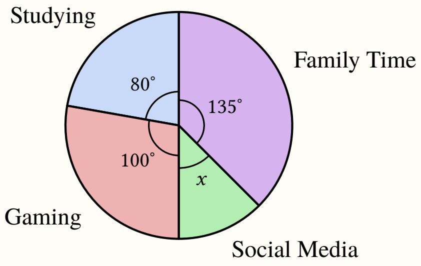
- Clearly shows how each category contributes to a whole, making it ideal for displaying proportions.
- Reduce misinterpretation by avoiding axes and scales.
- More visually engaging.

Cons

- Curved areas and angles are difficult to interpret precisely, especially for small differences.
- Suffers from cluttering with many categories.
- Focuses on relative proportions instead on exact values.

Example 1:

The pie chart shows how Yuze spent her 20 waking hours on Saturday.



a) Find the value of x.
b) Thus, find the number of hours Yuze spent on Social Media.
c) What percentage of Yuze's waking hours was spent with family?
d) If Yuze wants to reduce her gaming time by half and use that time to study instead, how many total hours would she spend studying?

a) – $x = 360^\circ - 135^\circ - 80^\circ - 100^\circ = 45^\circ$

b) – $\frac{45^\circ}{360^\circ} \times 20 = 2.5$ hours
– Yuze spent 2.5 hours on Social Media.

c) – $\frac{135^\circ}{360^\circ} \times 100\% = 37.5\%$
– Yuze spent 37.5% of her waking hours with family.

d) – $\frac{80^\circ + \frac{100^\circ}{2}}{360^\circ} \times 20 = 7.22$ hours
– She would spend 7.22 hours on studying.

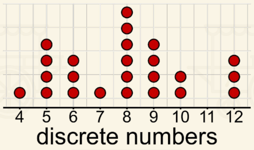
2

Discrete Numbers

Numbers can be discrete too!

Dot Diagram

A dot diagram shows individual data points placed along a number line. It is great for small sets of discrete numerical data and helps show counts and clustering clearly.



Pros

- Displays exact data points while also making patterns easy to identify.
- Easy to construct and interpret.

Cons

- Hard to read for large datasets.
- Only works for discrete numbers, not effective for continuous numbers.
- Not effective with many unique data points or wide ranging data.

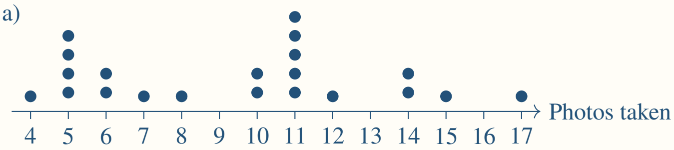
Example 1:

Dorothy went on a 21-day-long odyssey to Japan. She records the number of photos she took every day during the trip in the table below.

6	14	5	11	11	10	12
10	15	11	5	4	8	11
11	5	5	6	17	7	14

a) Create a dot diagram based on the data above.
b) What is the range of the number of photos taken?
c) What is the mean number of photos taken per day?
d) Find the mode. Hence, find the number of days where she took fewer photos than the mode.
e) Briefly describe the distribution of the data.

a)



b) – Range $= 17 - 4 = 13$

c) – Total $= 4 + 5 \times 4 + 6 \times 2 + 7 + 8 + 10 \times 2$
$+ 11 \times 5 + 12 + 14 \times 2 + 15 + 17$
$= 198$ photos

– Mean $= \frac{198}{21} = 9.43$ photos

d) – Mode $= 11$
– $1 + 4 + 2 + 1 + 1 + 2 = 11$
– There were 11 days where she took fewer than 11 photos a day.

e) – The number of photos of the 21 days range from 4 to 17. It is clustered around 5 and 11 photos a day. The distribution is not symmetrical.

Stem-and-leaf Diagram

A stem-and-leaf diagram splits each number into a “stem” and a “leaf” to organize discrete numerical data while preserving all individual values. It groups numbers by leading digits, allowing for a compact display of a wide range of values.

5	3 5 8 9 9
6	0 0 1 4 4 5 7 7
7	0 3 6 6 7
8	1 2 6 7
9	0 1

Key: 5 | 3 represents 53.

Pros (vs Dot Diagram)

- Groups numbers into bins, making it easier to represent wide ranging data.

Cons

- Binning is limited by the number structure.
- Requires a key for interpretation.
- Hard to read for large datasets.
- Less intuitive for the general audience.

Back-to-back Stem-and-leaf Diagram

Comparing two datasets using the same stems shows both distributions in a compact format.

0 3 3 4 6 8 8	3	
0 0 1 2 6 7	4	
2 4 5 8	5	3 5 8 9 9
1 3 7 8 9	6	0 0 1 4 4 5 7 7
0 0	7	0 3 6 6 7
	8	1 2 6 7
	9	0 1

Key: 0 | 3 represents 30. Key: 5 | 3 represents 53.

Example 1:

A PE teacher recorded the number of push-ups male and female students could complete in a minute. The results are displayed in the tables.

Male Students

24	31	45	38	52	39	36	40
33	47	28	22	55	34	19	41

Female Students

12	8	18	24	15	20	6	10
17	13	5	22	16	9	19	11

a) Create a back-to-back stem-and-leaf diagram based on the data.
b) What is the median number of push-ups for each group?
c) Find the range of each group.
d) Which interval was the most common regardless of group?

a)

	0	5 6 8 9
9	1	0 2 3 5 6 7 8 9 11
8 4 2	2	0 2 4
9 8 6 4 3 1	3	
7 5 1 0	4	
5 2	5	

9 | 1 represents 19. 0 | 5 represents 5.

b)
- $\text{Median}_{\text{M}} = \dfrac{36+38}{2} = 37$ push-ups
- $\text{Median}_{\text{F}} = \dfrac{15+16}{2} = 15.5$ push-ups

c)
- $\text{Range}_{\text{M}} = 55 - 19 = 36$
- $\text{Range}_{\text{F}} = 24 - 5 = 19$

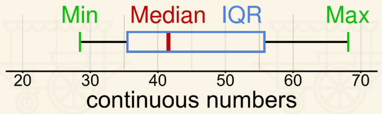
d)
- The most common interval was 10 to 19 push-ups with 10 students.

3 Continuous Numbers

Visualising continuous data is trickier because it can take on infinitely precise values and needs to be grouped before patterns emerge.

Box-and-whisker Plot

A box-and-whisker plot shows the spread and symmetry of data using five key values: minimum, lower quartile, median, upper quartile, and maximum.



Pros

- Summarises data using quartiles, giving a quick overview of spread and center.
- Space efficient and easy to construct with quartiles computed.
- Handles large datasets more efficiently, without becoming cluttered.

Cons

- Does not show shape of distribution such as peaks or gaps.
- Can be misleading if data is uneven or multimodal (multiple peaks).
- Exact data values are hidden in intervals.

Example 1:

Students recorded their weekly screen time (in hours). The results are displayed in the tables.

Lower Secondary

18.9	9.3	10.9	12.4	11.3	13.2	14.9	15.6	7.7	12.0

Upper Secondary

18.9	14.8	19.0	16.0	18.5	15.9	21.1	23.3	26.4	12.6

a) Create two box-and-whisker plots based on the data.
b) Which group has the higher median screen time?
c) Which group has the wider interquartile range?
d) Which group has the larger range?

a) Lower Secondary

7.7 9.3 10.9 11.3 12.0 12.4 13.2 14.9 15.6 18.9

Q_1 (10.9) Q_2 Q_3 (14.9)

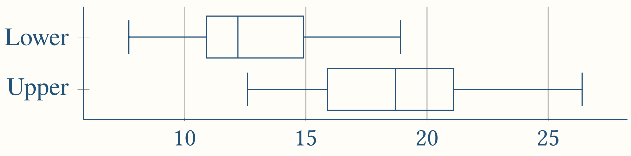
- $Q_2 = \dfrac{12.0 + 12.4}{2} = 12.2$

Upper Secondary

12.6 14.8 15.9 16.0 18.5 18.9 19.0 21.1 23.3 26.4

Q_1 (15.9) Q_2 Q_3 (21.1)

- $Q_2 \dfrac{18.5 + 18.9}{2} = 18.7$



b)
- Upper Secondary students have the higher median screen time.

c)
- $\text{IQR}_L = 14.9 - 10.9 = 4.0$
- $\text{IQR}_U = 21.1 - 15.9 = 5.2$
- Upper Secondary students have the higher interquartile range.

d)
- $\text{Range}_L = 18.9 - 7.7 = 11.2$
- $\text{Range}_U = 26.4 - 12.6 = 13.8$
- Upper Secondary students have the larger range of screen time.

Interval Frequency Table

Continuous data can vary widely and generally do not repeat exactly. Instead of listing every number, we can group them into intervals to more effectively determine distribution.

Raw Data

52.32	32.76	33.24	38.23
41.12	42.00	45.79	54.22
64.67	54.94	68.15	28.47
58.16	34.16	38.06	67.15
60.97	47.20	35.95	63.29
39.77	29.65	39.51	30.42

Table

Intervals	Counts
$20 \le x \le 30$	2
$30 < x \le 40$	9
$40 < x \le 50$	4
$50 < x \le 60$	4
$60 < x \le 70$	5

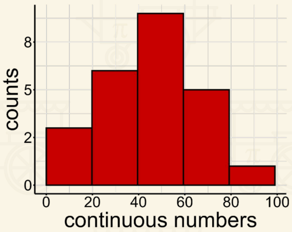
Histogram

A histogram displays the distribution of continuous data using touching vertical bars. Each bar represents an equal-width interval arranged in order. It can be directly constructed using an interval frequency table.

Table

Intervals	Counts
$20 \le x \le 30$	3
$30 < x \le 40$	6
$40 < x \le 50$	9
$50 < x \le 60$	5
$60 < x \le 70$	1

Diagram



Pros (vs Box-and-whisker)

- Shows approximate shape of distribution such as peaks or gaps.
- Flexible bin sizes allow for custom grouping based on context or precision.

Cons

- Choice of bin size can affect interpretation, potentially hiding or exaggerating patterns.
- Exact data values are hidden in intervals.
- Less intuitive for general audience and easily confused with bar graphs.

Example 1:

A teacher recorded the typing speeds of 20 students during an assessment. The results are displayed in the table in words per minute (wpm).

54	71	33	53	49	28	81	70	52	74
45	36	58	64	41	54	62	67	44	35

a) Create an interval frequency table based on the data.

b) Hence, create a histogram based on the data.

c) Briefly describe the distribution of the data.

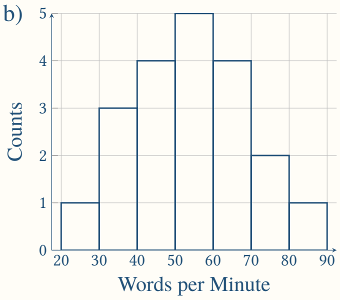
d) Find the percentage of students in the most common interval.

e) After more students were tested, the 40-50 wpm interval becomes most common. Find the minimum number of students tested.

a)

Intervals	Counts
$20 \le x \le 30$	1
$30 < x \le 40$	3
$40 < x \le 50$	4
$50 < x \le 60$	5
$60 < x \le 70$	4
$70 < x \le 80$	2
$80 < x \le 90$	1

b)



c) – The typing speed ranges from 20 to 80. Typing speed is most common in and roughly symmetric about the 50-60 words per minute interval.

d) – $\frac{5}{20} \times 100\% = 25\%$

e) – $5 - 4 + 1 = 2$

– At least 2 students were tested.

Cumulative Frequency Table

A cumulative frequency table shows the running total of frequencies as values increase. It helps track how many data points fall below or at certain values.

Interval Table

Intervals	Counts
$20 \le x \le 30$	3
$30 < x \le 40$	6
$40 < x \le 50$	9
$50 < x \le 60$	5
$60 < x \le 70$	1

Cumulative Table

Intervals	Counts
$x \le 20$	0
$x \le 30$	3
$x \le 40$	9
$x \le 50$	18
$x \le 60$	23
$x \le 70$	24

+3
+6
+9
+5
+1

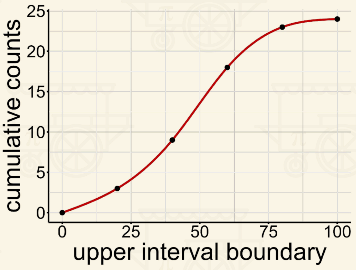
Cumulative Frequency Curve

A cumulative frequency curve is a smooth graph that shows how data accumulates over time or across values. It's useful for estimating medians, quartiles, and understanding the overall distribution of data. It can be directly constructed using a cumulative frequency table.

Table

Intervals	Counts
$x \le 20$	0
$x \le 30$	3
$x \le 40$	9
$x \le 50$	18
$x \le 60$	23
$x \le 70$	24

Diagram



Pros (vs Histogram)

- Easier to estimate key statistics like the median, quartiles, and percentiles.
- Smooth curve makes trends clearer, especially for identifying skewness or spread.

Cons

- Difficult to identify modes and peaks.
- Exact data values are hidden in intervals.
- Less intuitive for general audience.

Example 1:

A teacher recorded the typing speeds of 20 students during an assessment. The results are displayed in the table in words per minute (wpm).

54	71	33	53	49	28	81	70	52	74
45	36	58	64	41	54	62	67	44	35

a) Create a cumulative frequency table based on the data.

b) Hence, create a cumulative frequency curve based on the data.

c) Estimate the number of students that type slower than 45 wpm.

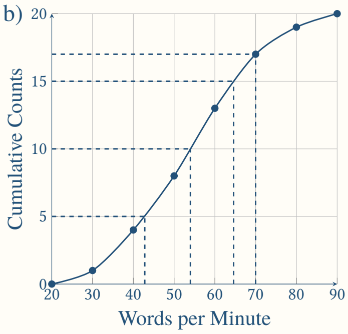
d) Estimate the quartiles Q_1, Q_2, and Q_3.

e) Estimate the typing speed to be in the top 15% of students.

a)

Intervals	Counts
$x \le 20$	0
$x \le 30$	1
$x \le 40$	4
$x \le 50$	8
$x \le 60$	13
$x \le 70$	17
$x \le 80$	19
$x \le 90$	20

b)



c) – From the graph, ~6 students type slower than 45 wpm.

d) – $\frac{1}{4} \times 20 = 5$ – From the graph, $Q_1 \approx 42$.

– $\frac{2}{4} \times 20 = 10$ – From the graph, $Q_1 \approx 55$.

– $\frac{3}{4} \times 20 = 15$ – From the graph, $Q_3 \approx 65$.

e) – $\frac{100-15}{100} \times 20 = 17$

– From the graph, a greater than ~70 wpm is required to be in the top 15%.

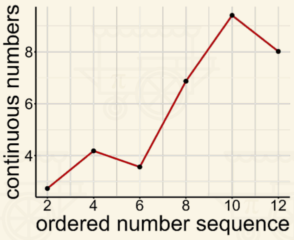
Line Graph

A line graph displays trends in numerical data over an ordered sequence (often time). Points are plotted and connected by lines to show trends in the data.

Raw Data

Ordered Sequence	2.0	4.0	6.0	8.0	10.0	12.0
Continuous Data	2.73	4.18	3.56	6.87	9.41	8.02

Diagram



Pros

- Easy to see increases, decreases, and patterns at a glance.
- Can display multiple datasets (multiple lines) on the same graph for easy comparison.
- Emphasizes rate of change as lines represent gradients.

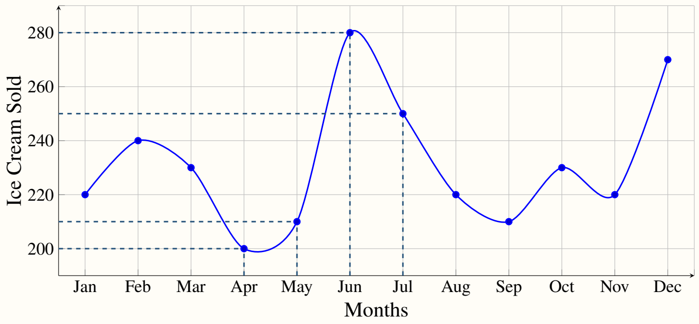
Cons

- Not suitable for categorical data, since the x-axis must follow a logical order.
- Assumes a straight-line (linear) change between points, which may not reflect real data.
- Less effective with small datasets, where a bar graph or dot plot may be clearer.
- Axis truncation can lead to misinterpretation.

Unlike the other graphs for continuous data, the line graph is the only 2-dimensional one in the syllabus. The others focus on visualizing the distribution of a single set of continuous data.

Example 1:

The line graph below shows the number of ice cream cones sold per month by a vendor across a year. Each ice cream was sold for $4.



a) In which months were the sales the highest and the lowest?
b) Which two consecutive months had the largest increase in sales?
c) Estimate the total revenue for the year.
d) Estimate percentage decrease in sales from June to July.

a) – From the graph, April had the lowest sales at 200 ice creams sold, while June had the highest sales at 280 ice creams sold.

b) – The largest increase in sales was between May and June from 210 to 280 ice creams sold.

c) –
$$\begin{aligned}\text{Ice Cream Sold} &= 220+240+230+200+210+280\\ &\quad+250+220+210+230+220+270\\ &= 2780\end{aligned}$$

– Revenue $= 2780 \times 4 = 11120$

d) – $\frac{280-250}{280} \times 100\% = 10.7\%$

Bar Graph vs Histogram

Although histograms and bar graphs look similar, they are used for different types of data and follow different rules.

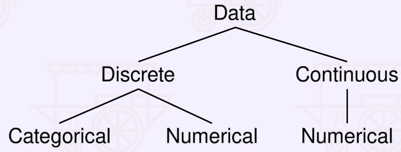
	Bar Graph	Histogram
Purpose	Compare counts of discrete values.	Show distribution of continuous data.
Graphic	Vertical bars with gaps.	Touching vertical bars.
Construction	Plot counts directly against corresponding discrete values.	Bin continuous data into equal intervals. Count values in bins. Plot count against intervals.

Data Types

Discrete data comes in separate values you can count. There are two types of discrete values:

- **Discrete categorical data** uses labels to represent types or groups. Examples: shirt sizes (S, M, L, XL), animal types (dog, cat, fish).
- **Discrete numerical data** can only take on whole number values with clear gaps between them. Examples: number of books (0, 3, 24), goals scored (0, 1, 2).

Continuous data can take any value within a range, including decimals. By definition, continuous data can only be numerical. Examples: heights of students (153.32, 161.42, 168.01), time taken per lap (14.35, 15.99, 18, 12).



REACH US AT:

MATHHAWKER.COM.SG – ALL NOTES

@MATHHAWKER
@MATHHAWKER – LATEST UPDATES

PATREON.COM/MATHHAWKER – SUPPORT US :)