

1 CHAPTER OVERVIEW

Your visual roadmap through the chapter

Data, Types & Collection

Classifying data and how it's gathered

1

Data Organization & Summary

Arrays, frequency distributions, tally, tabulation

2

Graphical Representation

Bar/pie charts, histograms, ogives, stem-leaf, time series

3

WHAT IT'S ABOUT

Data is the source of information. This chapter covers the different types of data, how it is collected, how raw data is organized and summarized, and how it is represented graphically.

WHY IT MATTERS FOR QAB

Recognising data types and choosing the right organizing/graphing method is the foundation for all later statistical analysis and interpretation topics.

EXAM SKILLS REQUIRED

Classify data correctly; build frequency distributions and tables; construct/read histograms, ogives and stem-leaf plots; identify fractiles from a graph.

2 KEY DEFINITIONS

Source-based terminology, exam-ready

Data vs Information

Data refers to facts, figures or values. Information is derived when facts are processed, structured and analysed for decision making.

Key word: raw vs processed

Numerical / Categorical

Numerical (quantitative) data allows arithmetic operations; categorical (qualitative) data does not.

Key word: can you calculate it?

Discrete / Continuous

Discrete variables take only certain, separate values (a count). Continuous variables can take any value in a range.

Key word: count vs measure

Ordinal / Nominal

Ordinal data has a natural order. Nominal data has no natural order — reordering doesn't change its meaning.

Key word: order matters or not

Cross-sectional / Time Series

Cross-sectional: data at one distinct point in time. Time series: observations of one phenomenon tracked over time.

Key word: one moment vs over time

Primary / Secondary Data

Primary data is collected first-hand for the investigation. Secondary data was previously collected and may be relevant.

Key word: first-hand vs reused

Population / Sample

Population data comes from investigating every member. Sample data is a representative selection within the population.

Key word: everyone vs a selection

Structured / Unstructured

Structured data has a pre-defined model or organization. Unstructured data has no organized form.

Key word: organized or not

3 CORE CONCEPT

Explained in exam-oriented language

CONCEPT → Organizing Raw Data

MEANING

Raw data is rearranged into arrays, frequency distributions and tables so it becomes easier to understand.

APPLICATION

Summarising large datasets (e.g. 100 salesmen's distances) before analysis.

PURPOSE

"The human mind cannot assimilate data" in its raw form — structure makes it usable.

CONCEPT → Graphical Representation

MEANING

Organized data is displayed visually — some methods simple (bar, pie), others more sophisticated (histogram, ogive).

APPLICATION

Choosing the right chart for categorical vs grouped numerical data.

PURPOSE

Show relationships, proportions and distributions at a glance.

4 KEY RULES & PRINCIPLES

Every source rule, exam-ready

RULE 1 Constructing grouped classes

- Identify the minimum and maximum term to create classes
- Keep the number of classes relatively small (6–7, or 15–20 for large datasets)
- Classes should generally be of equal width for easier comparison
- Class limits must be unambiguous (e.g. "0 to less than 10")
- The upper limit of one class is the lower limit of the next — no gaps

RULE 3 Principles of data tabulation

- Simplicity: classify and keep details to a minimum
- Title: a clear title explaining the table's purpose
- Source, Units and Headings must be clearly labelled
- Totals and subtotals shown where meaningful

RULE 2 Data quality

- Useful data should be: complete, relevant & significant, timely, appropriately detailed, clear, properly communicated, and reliable/consistent

RULE 4 Unequal class widths in a histogram

- Either calculate frequency density (frequency ÷ class width), or
- Adjust the frequency to what it would be for the standard width, then use that as the bar height

5 FORMULA ZONE

Every formula supported by the source text

RELATIVE FREQUENCY

Relative Frequency = $\text{Class Frequency} \div \text{Total Frequency}$

The sum of all relative frequencies in a distribution always equals 1.

FREQUENCY DENSITY

Frequency Density = $\text{Frequency} \div \text{Class Width}$

Used to build a histogram bar height when class widths are unequal.

CLASS WIDTH, BOUNDARY & MID-POINT

Width = $\text{Range} \div \text{No. of Classes}$

Class boundary: midpoint between one class's upper limit and the next class's lower limit (± 0.5 for discrete data). Mid-point = $(\text{upper limit} + \text{lower limit}) \div 2$.

FRACTILES (PERCENTILE / DECILE / QUARTILE)

p^{th} percentile: $p\%$ of items fall at or below it

Decile = $1/10$ of the way through; Quartile = $1/4$ of the way through. Q_2 = 50th percentile = the median.

6 CALCULATION ROADMAP

General path for any data-organization question

1- Read Question

Identify the raw data and what's required

2- Identify Method

Array, frequency table, chart, or fractile?

3- Organize

Sort, tally, or group into classes

4- Calculate/Plot

Apply the formula, or plot the chart

5- Interpret

Read off / state the final answer

7 STEP-BY-STEP METHODS

Kept separate — do not mix methods

METHOD A · Building a Grouped Frequency Distribution

- 1 Find the minimum and maximum values (the range)
- 2 Decide the number of classes; divide range by classes for width
- 3 Set unambiguous class limits (e.g. "480 to under 500")
- 4 Tally each data value into its class, then count for frequency

METHOD C · Histogram with Unequal Class Widths

- 1 Note each class's frequency and width
- 2 Calculate frequency density = frequency ÷ class width for each class
- 3 Plot frequency density as the bar height (not raw frequency)
- 4 Alternative: adjust frequency to a common width, then use as height

METHOD B · Cumulative Frequency (Less/More Than)

- 1 List classes in order with their individual frequencies
- 2 "Less than" series: add each class's frequency to the running total of all previous classes
- 3 "More than" series: start at the total and subtract each class's frequency going forward
- 4 Use for plotting an ogive

METHOD D · Reading a Fractile from an Ogive

- 1 Find total number of items (n) in the distribution
- 2 Calculate the position: e.g. Q1 is the 25% × n th item
- 3 Find that position on the vertical (cumulative frequency) axis
- 4 Read across to the curve, then down to the horizontal axis for the value

METHOD E · Constructing a Stem-and-Leaf Plot

- 1 Sort the data into ascending order
- 2 Split each value into a stem (leading digits) and a leaf (last digit)
- 3 List stems once down the left; write each value's leaf beside its stem
- 4 Define the units for stem and leaf clearly (e.g. stem=10s, leaf=1s)

8 DIFFERENT CASES / SCENARIOS

Recognise the data, pick the right chart or method

IF data is categorical (e.g. proportions of a whole)

→ USE a **bar chart** or **pie chart**

IF data is a grouped frequency distribution

→ USE a **histogram** (density if class widths differ)

IF you need a running/cumulative total plotted

→ USE an **ogive**

IF you need to keep the actual raw data values visible

→ USE a **stem-and-leaf plot**

IF you need to show a variable's trend over time

→ USE a **time series graph**

IF class widths in a histogram are NOT equal

→ USE **frequency density**, not raw frequency, as bar height

IF asked for a percentile/decile/quartile value

→ Find its position, then read it off an **ogive**

IF data was collected specifically for this investigation

→ It is **primary** data; if reused from elsewhere, it's **secondary**

9 COMPARISON TABLES

Don't confuse these!

	Numerical (Quantitative)	Categorical (Qualitative)
Arithmetic possible?	Yes	No
Examples	Age, salary, temperature	Eye colour, location, gender
	Discrete	Continuous
Values	Certain, separate values (a count)	Any value within a range
Examples	Number of contracts, children	Height, weight
	Ordinal	Nominal
Order	Has a natural order	No natural order
Examples	A/B/C grade, small-medium-large	Gender, pass-fail
	"Less Than" Cumulative Series	"More Than" Cumulative Series
Direction	Adds frequencies going up through classes	Subtracts frequencies going up through classes
Starts at	0	Total frequency

10 GRAPH & DATA INTERPRETATION

How to read the source's charts and tables

Reading an ogive

Each point shows how many items fall below the corresponding value on the horizontal axis (a "less than" ogive). To find a fractile, locate its position on the vertical axis, trace across to the curve, then down to read the value.

Reading a histogram

Bar height = frequency when class widths are equal, or frequency density when they are not. There are no gaps between bars, since the data is continuous/grouped.

Reading a stem-and-leaf plot

The stem (left of the line) holds the leading digit(s); each leaf (right of the line) is one data point's last digit. Reading a stem with all its leaves reconstructs every original value in that class.

Reading pie & component bar charts

A pie chart shows each category's share of the whole (slice size = proportion/percentage). A percentage component bar chart shows every bar at equal height, split into segments by percentage share instead.

11 WORKED CALCULATION LOGIC

Method students must reproduce in the exam

GIVEN → FORMULA → SUBSTITUTION → CALCULATION → ANSWER

An ogive shows a distribution of 40 items. What is the 1st quartile of the distribution?

GIVEN: $n = 40$ items; $Q1$ = the item 25% of the way through

FORMULA: Position of $Q1 = 25\% \times n$

SUBSTITUTION: Position = 0.25×40

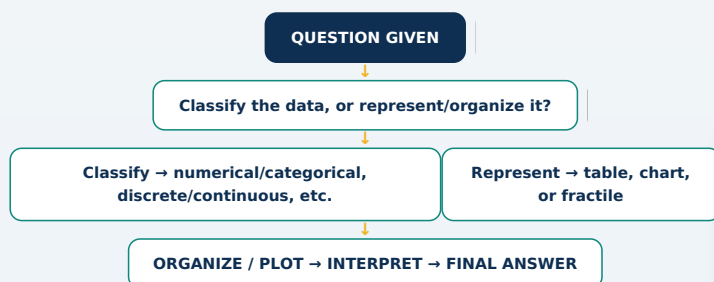
CALCULATION: Position = the 10th item — locate 10 on the vertical (cumulative frequency) axis, trace to the curve, then down to the horizontal axis

FINAL ANSWER: Reading off the ogive at the 10th item gives a 1st quartile of approximately 110

12 EXAM QUESTION PATTERNS

- **Classification** — identify data as numerical/categorical, discrete/continuous, etc.
- **Frequency distribution** — build a grouped table from raw data
- **Chart selection** — choose the appropriate chart for given data
- **Fractile reading** — find a percentile/decile/quartile from an ogive
- **Density calculation** — histogram with unequal class widths
- **Interpretation** — reading stem-leaf, pie, or component bar charts

14 QUICK DECISION TREE



13 EXAM TRAPS

⚠ WATCH OUT

- ✗ **Discrete vs continuous confusion** — a count (discrete) is not the same as a measurement (continuous).
- ✗ **Using raw frequency as histogram height** when class widths are unequal — use frequency density instead.
- ✗ **Ambiguous class limits** — always state classes as "X to under Y" to avoid double-counting a boundary value.
- ✗ **Less than vs more than series mix-up** — one accumulates upward, the other counts down from the total.
- ✗ **Misreading fractile position** — e.g. using n instead of the correct fraction of n (25% for $Q1$).

CHAPTER 7 IN 60 SECONDS

Data types: numeric/categorical, discrete/continuous, ordinal/nominal

Primary = first-hand; **secondary** = reused

Relative freq = $\text{class } f \div \text{total } f$

Frequency density = $f \div \text{class width}$

Charts: bar/pie (categorical), histogram/ogive (grouped)

Ogive = cumulative frequency curve

Q2 = median = 50th percentile

Trap: unequal widths need frequency density