



Analysing scatter plots (TARSOG)

Key focus

This objective introduces the TARSOG framework for describing a bivariate scatter plot: *Trend* (linear or non-linear), *Association* (positive or negative), *Relationship* strength (weak, moderate, or strong), *Scatter* (constant or non-constant spread), *Outliers*, and *Groupings*. Unlike time series trend (which must be “increasing”/“decreasing”), association in a scatter plot is correctly described as *positive* or *negative* – these are two different frameworks describing two different kinds of graph, and students should not mix the vocabulary up. Students learn to read each component off a plot individually, then combine all six into a full description, and to reason about how one component (an outlier, a grouping) can change how the others should be read.

Suggested teaching sequence

1. Introduce the acronym TARSOG and define each letter before showing any plots: *Trend* (linear/non-linear), *Association* (positive/negative), *Relationship* (weak/moderate/strong), *Scatter* (constant/non-constant), *Outlier*, *Grouping*.
2. Show a strong positive linear scatter plot (e.g. revision hours vs. test score). Model the description: “linear trend, positive association, strong relationship” – the points sit close to a straight, rising line.
3. Show a strong negative linear scatter plot (e.g. TV hours vs. test score) and contrast it directly with the positive example.
4. Show a non-linear plot (e.g. a curved, accelerating pattern). Discuss why a straight fit line would be misleading here, and how “non-linear” still allows a clear positive or negative overall direction.
5. Introduce weak vs. strong relationships by comparing two plots with the same overall direction but very different scatter around the trend – the strength describes how closely points cluster around the pattern, not the steepness of the line.
6. Introduce outliers: a point that sits well away from the overall pattern. Discuss how a single outlier can distort a visual impression of the trend or relationship strength if not identified separately.
7. Introduce groupings: two or more separate clusters of points, which often suggest a hidden third variable (e.g. two age groups, two genders, two conditions) rather than one single relationship.
8. Introduce constant vs. non-constant scatter: whether the spread of points around the trend stays roughly even, or changes (e.g. a funnel shape that widens) as the x-values increase.
9. Practise describing full, unfamiliar plots using all six TARSOG components together, justifying each choice with reference to what is actually visible in the plot.

Core examples to model

- Strong positive linear: revision hours vs. test score, points close to a rising straight line – “linear, positive, strong.”

- Strong negative linear: TV hours vs. test score, points close to a falling straight line – “linear, negative, strong.”
- Non-linear: a curve that accelerates upward – trend is still positive overall, but the pattern is not a straight line.
- Outlier: one point far from an otherwise clear positive linear pattern – describe the main pattern first, then note the outlier separately rather than letting it distort the description.
- Grouping: two visually separate clusters of points – describe each cluster’s own pattern if they differ, and note that a third variable (not shown on the axes) may explain the split.

Watch for

- Confusing scatter-plot association with time-series trend vocabulary: association here is correctly “positive”/“negative” – do not import the “increasing”/“decreasing” rule from time series analysis.
- Confusing relationship strength with trend steepness: a steep line with points scattered widely around it is still a *weak* relationship; strength is about how closely points sit to the pattern, not how steep it is.
- Treating an outlier as part of the main pattern: an outlier should be identified and set aside, not allowed to change the overall trend/association/strength description.
- Missing a grouping and describing two separate clusters as one single (often “weak” or “non-linear”) relationship, when in fact two clear linear relationships exist within two separate groups.
- Forcing a linear fit line onto an obviously non-linear or grouped plot: a straight line is only appropriate when the trend is genuinely linear.

Checks for understanding

- Can students state what each letter of TARSOG stands for without prompting?
- Can they correctly classify trend as linear or non-linear from a plot?
- Can they correctly classify association as positive or negative (and distinguish this from time-series trend direction)?
- Can they judge relationship strength (weak/moderate/strong) based on how closely points cluster, independent of the line’s steepness?
- Can they identify outliers and explain how they should be treated separately from the main pattern?
- Can they identify groupings and suggest what hidden variable might explain them?
- Can they identify constant vs. non-constant scatter?
- Can they combine all six components into one coherent description of an unfamiliar plot?