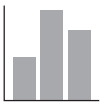
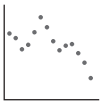
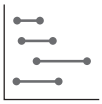
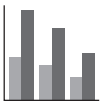
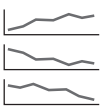
What type of graph is best for my data?

Graphs are the best way of displaying data when it is more important to convey overall patterns in the data than the individual data values. The overall shape of the graph reveals trends, differences, exceptions, anomalies and other relationships that might be difficult to detect in a table.

It is critical to choose the correct type of graph for your data so that these relationships are clear. More often than not, the simplest graphing option will be best. Readers will engage with, and be persuaded by, the message of your data if they are presented in a way that matches readers' intuitive understanding of data relationships – for example, that horizontal lines represent measurements over time. Some data relationships can be shown by more than 1 type of graph. In this situation, you should use the graph type that is most familiar to your particular readership.

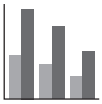
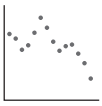
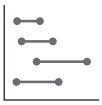
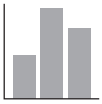
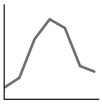
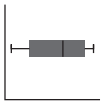
Type of data and relationship	Recommended graph type	Defining features
Change over time Shows how data values for a measure(s) change over time (eg population-adjusted breast cancer diagnoses recorded in Australia every year, for the past 20 years)	 Line graph (for large time series)	Use to highlight trends or patterns in a measure over many time points Use for datasets that include data for more than about 8 time points Lines are connected, consecutive data values Lines always follow a horizontal direction, with time intervals on the *x* axis increasing from left to right, and the measurement variable plotted on the *y* axis Only connect consecutive values – intervals with missing data must be shown as a break in the line
	 Vertical bar graph (for small time series)	Use for time-series data with a small number of time points (around 8 or less) Use to emphasise specific data values, rather than an overall pattern or trend
	 Dot graph	Dots represent data values at each time point. If connected, these dots form a line graph Can be mistaken for scatter plots – consider using a bar graph or line graph instead
	 Dumbbell graph	Connected dots represent 2 time points of data (eg pre- and post-test) for multiple groups
	 Vertical bar cluster graph	Can be used to show data for 2 or 3 time points of data with groups Can be difficult for readers to compare differences across many groups because of distance between groups on x axis
	 Trellis of single-category graphs	Consider using a trellis (panel) graph with smaller graphs for each group if you have a large number of groups or categories of data that would appear cluttered on a single line or bar graph Keep the axis range consistent across all graphs to enable comparisons

Australian manual of style is published by

Biotext

MACQUARIE University

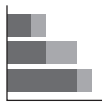
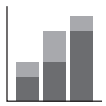
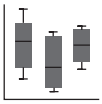
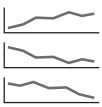
amos@stylemanual.com.au
02 6282 2280

Type of data and relationship	Recommended graph type	Defining features
Comparing groups or categories Compares data values across independent items, groups or categories (eg unemployment rates for each Australian state and territory)	 Horizontal bar graph	Order bars by size of data values to emphasise differences
	 Horizontal or vertical bar cluster graph	Use clustered bars for subcategories of groups, but limit clusters to 3 or 4 subcategories to enable comparisons across groups
	 Dot graph	Dots represent single data values for each item or group; a column of dots can represent summary values for each group Can be mistaken for scatter plots or time-series graphs – consider using a bar graph instead
	 Dumbbell graph	Connected dots represent 2 related data points on a common scale, for multiple groups Readers can easily judge distance between dots, and differences in this distance across groups
Single frequency or distribution data Shows how frequency or count values are distributed over the range of a measure (eg range of blood pressure measurements for men)	 Histogram (for measures with a small range)	Use a histogram to show frequency or count values across the range of a measure with few intervals Use as an alternative to a frequency polygon when individual data values must be emphasised
	 Frequency polygon (for measures with a large range)	Use to show frequency or count values across the range of a measure with many intervals Use to emphasise the shape of a distribution
	 Strip plot	Use to show the distribution of a measure for a small population If multiple measurements are recorded for the same value on the distribution, these points should be stacked or shown in a denser tone than other (nonrepeated) points
	 Box plot (horizontal or vertical)	Use to summarise a measure's distribution, rather than all individual data values May be unfamiliar to readers – consider plotting a simple histogram instead

Australian manual of style is published by

Biotext MACQUARIE University

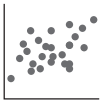
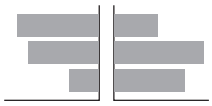
amos@stylemanual.com.au
02 6282 2280

Type of data and relationship	Recommended graph type	Defining features
Part to whole (ie proportions of a total) Shows how data values relate to, compare with, or make up a total measure at 1 or more points in time (eg proportion of Australia's total primary energy supply attributable to each major fuel type)	 Horizontal bar graph	Use to show the value (ie percentage or proportion of an absolute total) of each part for a single population This type of data is often shown as a pie graph, which is NOT recommended for displaying scientific data
	 Horizontal stacked bar graph	Use to show proportions of a total measure for multiple populations or groups Total(s) must add to 100% if parts are percentages, or the total absolute value for other measures
	 Vertical stacked bar graph	Use to show proportions of a total measure over time, for about 8 or fewer time points Use to emphasise changes in the relative size of parts over time
	 Stacked area graph	Use to show proportions of a total measure over time, for around 8 or more time points Use to emphasise changes in the relative size of parts over time
Distribution of the same measure across multiple time points or categories Shows how frequency or count values are distributed over the range of a measure, for more than 1 population (eg range of blood pressure measurements for men with 5 different medical conditions)	 Line graph with upper and lower bounds	Use to show distributions with a large number of time points – not multiple, discrete populations Summary values (eg median or mean values) for the distributions at each time point are connected to form a line The range (eg upper and lower confidence intervals) for the distribution at each time point are connected to form (typically invisible) lines above and below this line – the areas between the line and these upper and lower bounds lines are shaded Upper and lower bounds may be an unfamiliar feature for readers – consider whether their inclusion adds meaning and whether this outweighs potential misperceptions among readers
	 Strip plot	Multiple distributions are plotted side by side against the same *y* axis White space should separate each distribution See points above for strip plots
	 Vertical box plot	Use to summarise multiple distributions of the same measure May be unfamiliar to readers – consider plotting summary values (eg medians of the distribution) as a bar graph for multiple groups or populations, or a line graph with or without upper and lower bounds for multiple distributions over time
	 Trellis of single-category graphs	Consider using a trellis (panel) graph with smaller graphs for each group if you have a large number of groups or categories of data that would appear cluttered on a single line or bar graph Keep the axis range consistent across all graphs to enable accurate comparisons

Australian manual of style is published by

Biotext MACQUARIE University

amos@stylemanual.com.au
02 6282 2280

Type of data and relationship	Recommended graph type	Defining features
Deviation Shows the difference between data values and a baseline, target or threshold (eg differences between actual rainfall and predicted or previous-year rainfall for each month of a year)	 Vertical bar graph with baseline	Use when the goal is to highlight deviations between measurements and some meaningful baseline or reference Bars (ie data values) above the reference or *x* axis indicate positive differences from the baseline; bars below indicate negative differences The *y* axis can measure absolute differences or percentage change between data values and the reference
	 Line graph with baseline	Use to show differences from a baseline or reference over time when the dataset includes data for more than about 8 time points See points above for line graphs
Correlated measures Shows an association between 2 measures or variables (eg children's age and height)	 Scatter plot	Each dot or data point represents a subject's measurement on *x* axis and *y* axis variables Use to show that data points form a meaningful shape that indicates the type (or lack) of association between 2 variables Consider including a trend line to highlight the type and strength of association Depending on the audience, readers may be unable to interpret scatter plots – consider whether side-by-side horizontal bar graphs would better communicate the association
	 Side-by-side horizontal bar graph	Use to show an association between 2 measures when scatter plots are unfamiliar to readers Most effective for showing linear associations Two aligned bar graphs display each subject's measurement on the first and second measures Order the bars by size on one of the graphs to emphasise the association between the 2 measures

Note: Aspects adapted from Few S (2012). Show me the numbers: designing tables and graphs to enlighten, 2nd edn, Analytics Press, Burlingame, California; and Evergreen SDH (2017). Effective data visualization: the right chart for the right data, SAGE Publications, Thousand Oaks, California.

Australian manual of style is published by
Biotext
MACQUARIE University

amos@stylemanual.com.au
02 6282 2280