Choosing the right graph for your data

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

Choosing the correct type of graph for your data makes sure these important relationships are clear. Usually, the simplest type of graph will be best.

Readers will engage with and understand the data if they are presented in a way that is intuitive to them. For example, left-to-right lines are commonly understood to represent measurements over time.

Some relationships can be shown by more than one type of graph. In this case, choose a graph that is most familiar to your readers.

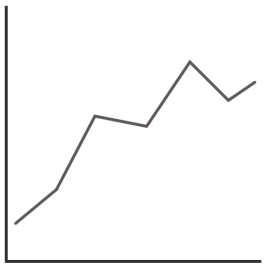
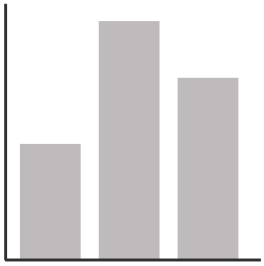
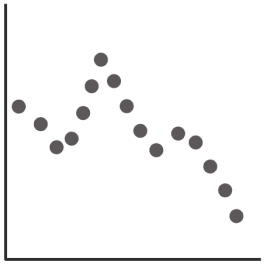
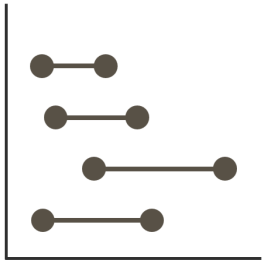
The following examples are adapted from Few (2012)[1] and give examples of graphs to show:

- change over time
- correlated variables
- comparing groups or categories
- single frequency or distribution data
- part to whole (or proportions of a whole)
- distribution of the same variable across many time points or categories
- deviation (from a baseline, target or threshold).

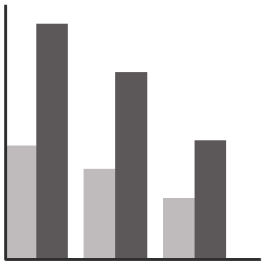
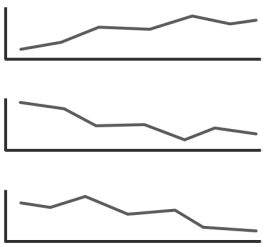
[1] Few S (2012). *Show me the numbers: designing tables and graphs to enlighten*, 2nd edn, Analytics Press, Burlingame, California; and Evergreen SDH (2017). *Effective data visualization: the right chart for the right data*, SAGE Publications, Thousand Oaks, California.

Change over time

These graphs show how data values for one or more variables change over time (e.g. population-adjusted breast cancer diagnoses recorded in Australia every year, for the past 20 years).

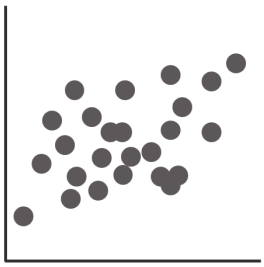
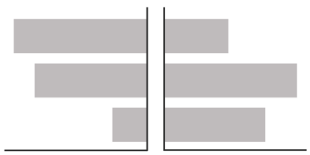
Graph type	Features and notes
 Line graph (for large time series)	• Use to highlight trends or patterns in a variable over many time points • Use for datasets that include data for more than about 8 time points • Lines connect consecutive data values • Lines always go left to right, with time intervals on the *x* axis increasing from left to right, and the variable plotted against the *y* axis • Only connect consecutive values – intervals with missing data must be shown as breaks in the line
 Vertical bar graph (for small time series)	• Use for time-series data with a small number of time points (about 8 or fewer) • Use to emphasise specific data values, rather than an overall pattern or trend
 Dot graph	• Dots represent data values at each time point • Can be mistaken for scatter plots – consider using a vertical bar graph or a line graph instead
 Dumbbell graph	• Connected dots represent 2 time points of data (e.g. pre- and post-test) for several groups • Use with care; this graph type is less common and will be unfamiliar to many readers

Change over time (continued)

Graph type	Features and notes
 Vertical bar cluster graph	• Can be used to show data for several groups at 2 or 3 time points • 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 a small graph for each group if you have many groups or categories of data and they would appear cluttered if they were all presented on the same line or bar graph • Keep the axis ranges consistent across all graphs to enable comparisons

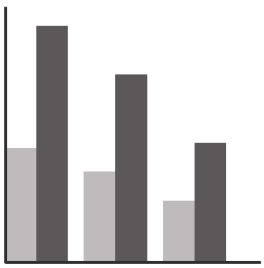
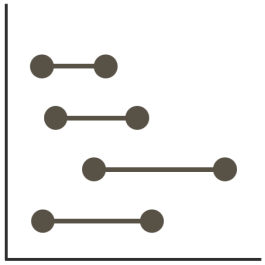
Correlated variables

These graphs show an association between 2 variables (e.g. children's ages and heights).

Graph type	Features and notes
 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 • Some readers may struggle to interpret scatter plots – consider whether side-by-side or stacked horizontal bar graphs would better communicate the association
 Side-by-side horizontal bar graph	• Use to show an association between 2 variables when scatter plots are unfamiliar to readers • Most effective for showing linear associations • Two aligned bar graphs display each subject's measurements on the first and second variables • Order the bars by size on 1 of the graphs to emphasise the association between the 2 variables

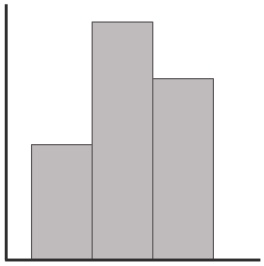
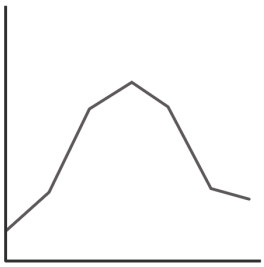
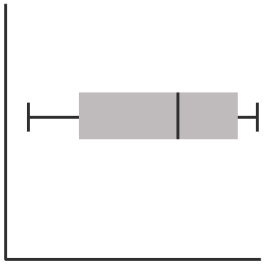
Comparing groups or categories

These graphs compare data values across independent items, groups or categories (e.g. unemployment rates for each Australian state and territory).

Graph type	Features and notes
 Horizontal bar graph	• If categories have no inherent order, can 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
 Dumbbell graph	• Connected dots represent 2 related data points on a common scale • Readers can easily judge distance between dots, and differences in this distance across groups • Use with care; this graph type is less common than the others and will be unfamiliar to many readers

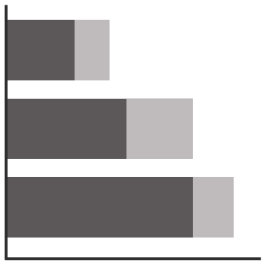
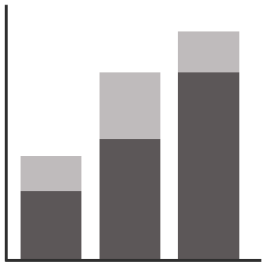
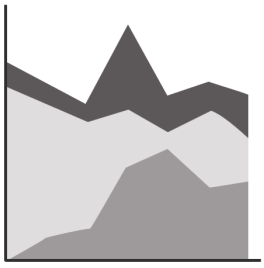
Single frequency or distribution data

These graphs show how possible values are distributed over the range of a variable (e.g. range of blood pressure measurements for people aged 60 to 70 years).

Graph type	Features and notes
 Histogram (for variables with a few possible discrete values)	• Use a histogram to show frequencies or count values across the range of a variable with few intervals • Use as an alternative to a frequency polygon when individual data values must be emphasised
 Frequency polygon (for variables with many possible values)	• Use to show frequency or count values across the range of a variable with many intervals • Use to emphasise the shape of a distribution
 Strip plot	• Use to show the distribution of a variable for a small population • If many measurements have the same value, these points should be stacked or shown in a denser tone than nonrepeated points • Use with care; this graph type is less common than some others and will be unfamiliar to many readers. Consider plotting a simple histogram instead
 Box plot (horizontal or vertical)	• Use to show a summary of a variable's distribution, rather than all individual data values • Use with care; this graph type is less common than some others and will be unfamiliar to many readers. Consider plotting a simple histogram instead

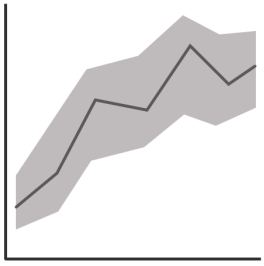
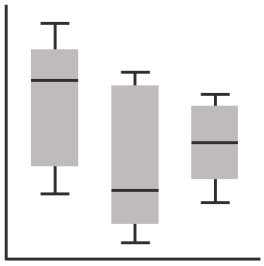
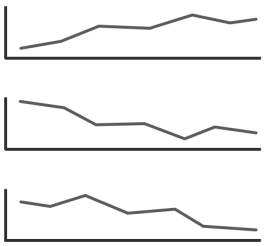
Part to whole (or proportions of a whole)

These graphs show how data values relate to, compare with or make up a total variable at one or more times, places and so on (e.g. proportion of Australia's total primary energy supply attributable to each major fuel type).

Graph type	Features and notes
 Horizontal bar graph	• Use to show the value (i.e. 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 variable for several populations or groups • Total(s) must add to 100% if parts are percentages, or to the total absolute value for other variables
 Vertical stacked bar graph	• Use to show proportions of a total variable for several times, populations or groups, for about 8 or fewer populations or groups • These groups may be the same group, measured at various times • Use to emphasise changes in the relative sizes of parts over time
 Stacked area graph	• Use to show proportions of a total variable over time, for about 8 or more time points • Use to emphasise changes in the relative sizes of parts over time

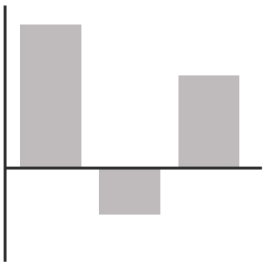
Distribution of the same variable across time points or categories

These graphs show how frequency or count values are distributed across the range of a variable, possibly for more than one population (e.g. range of blood pressure measurements for people aged 60 to 70 years with 5 different medical conditions).

Graph type	Features and notes
 Line graph with upper and lower bounds	• Use to show the distribution of a variable as it varies with time for many time points • Summary values (e.g. median or mean values) for the variable at each time point are connected to form a line • The upper confidence limits at each time are connected to form a line above the data, the lower confident limits are connected to form a line below the data, and the area between these lines is shaded • Upper and lower bounds may be unfamiliar to readers – consider whether their inclusion adds meaning and whether this outweighs potential misunderstandings
 Strip plot	• Distributions are plotted side-by-side against the same *y* axis • Use white space to separate the distributions • Use to show the distribution of a variable for a small population for a few categories or times on the *x* axis • If several measurements have the same value for a given distribution, these points should be stacked (for example, moved sideways by differing amounts) or shown in a denser tone than nonrepeated points
 Vertical box plot	• Use to summarise multiple distributions of the same variable, perhaps over time or across categories • May be unfamiliar to readers – consider plotting summary values (e.g. medians) 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 a small graph for each group if you have many groups or categories of data and they would appear cluttered if they were all presented on the same line or bar graph • Keep the axis ranges consistent across all graphs to enable accurate comparisons

Deviation

These graphs show the difference between data values and a baseline, target or threshold (e.g. differences between actual rainfall and predicted or previous-year rainfall for each month of a year).

Graph type	Features and notes
 Vertical bar graph with baseline	• Use when the goal is to highlight deviations of measurements from some meaningful baseline or reference • Bars (i.e. 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 • Use to highlight trends in a variable over many time points • Lines are connected, consecutive data values • Lines always go left to right, with time on the *x* axis increasing from left to right, and the measurement variable plotted against the *y* axis • Only connect consecutive values – missing data must be shown by breaks in the line