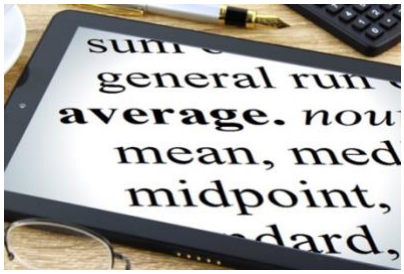


1

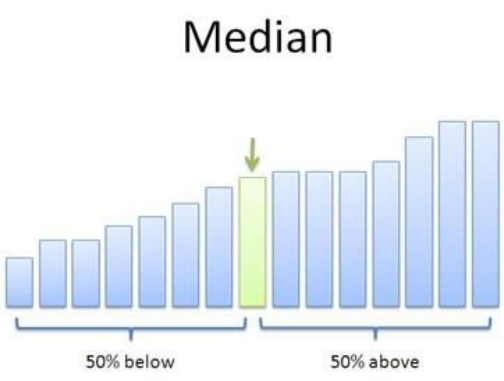
Descriptive Statistics

Descriptive statistics are used to summarise your data. These may be all you need but, even if you plan to run some statistical tests, you would usually start with an overview of your data. You might include a table showing the numbers in different categories, calculate some figures which are representative of your data and illustrate your data using graphs and charts.

Averages

An **average (Measure of Central Tendency)** is a number which can be thought of as being a 'middle' or 'typical' value of a variable. The average you use will depend on the kind of data you have.

The **mean**, often referred to as the average (for instance on Excel), is suitable for scale data.
It is calculated by adding up the measurements for all of your participants (readings) and dividing by the number of participants (readings). One way to think of this is that all the measurements are pooled and then shared out equally among the participants. For instance, the mean wage for a firm would be what would be paid to each worker if the total wage bill was divided equally among them.
Note that the mean can be affected by extreme values (outliers) or skewed data. For instance, one very highly-paid worker would increase the mean wage.
You might use the mean to compare different groups. For instance, you could compare the mean test marks in different classes in a school.
Sometimes the mean is used for ordinal data – but you should check what is expected in your course.



(2)

The **median** is the middle value of the data when it is arranged in order. It can be used for scale or ordinal data. It has the advantage that it is not affected by extreme values or skewed data. For instance, 'average' salary usually refers to the median as this will not be affected by a few people with high earnings.

The **mode** is the category or measurement with the highest frequency. It is the only average that can be used with categorical data. You should look at your data to see whether this is a useful average to report.

Measures of Spread (Dispersion)

Measures of spread (dispersion) can be used in conjunction with an average to provide further information about the shape of your data.

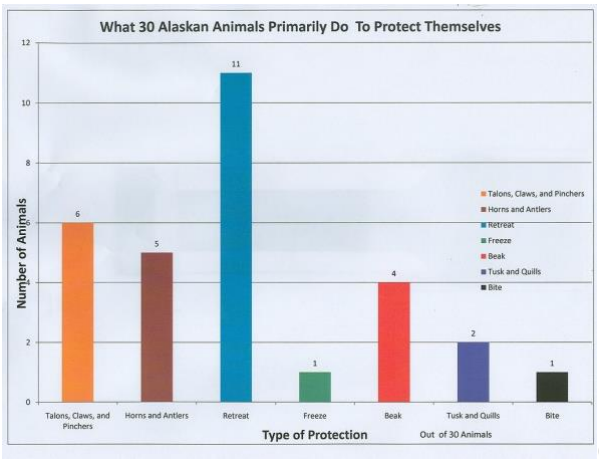
Quartiles are used with the median. There are different formulas for calculating their value so you should check which one you are expected to use. However, in general quartiles split the data for a variable into four equal parts. One quarter of your sample will have a value above the *upper quartile*, and one quarter will have a value below the *lower quartile*. The *interquartile range* is the difference between the upper and lower quartiles and so tells you the range in which half of the data sits. Quartiles are used in box plots.

The **standard deviation** is used with the mean for scale data. It gives information about how dispersed from the mean your values are. The number on its own needs to be considered in terms of the measurement you are interested in. It can be useful in comparing the same measurement in different groups.

Graphs, Tables and Charts

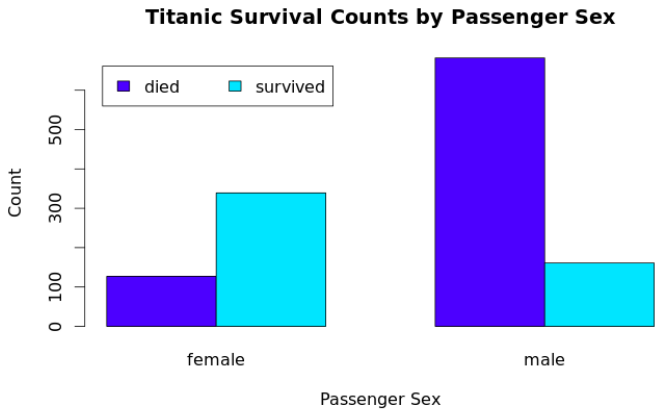
Graphs, tables and **charts** can be used to illustrate your data. You need to consider what is most appropriate for the kinds of data you have.
You will probably use a computer program (e.g. Excel, Minitab or SPSS) to draw the graphs for you. Avoid three-dimensional options as these can distort the values you are trying to show.

A **bar chart** can be used for categorical data.



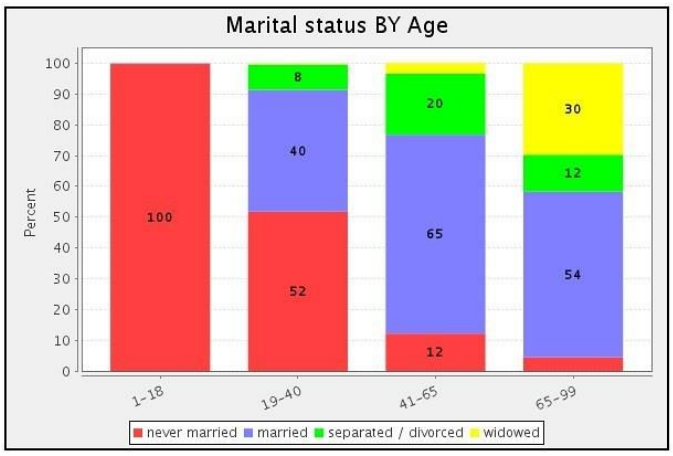
(3)

Clustered bar charts can be used to show two different variables on the same chart.



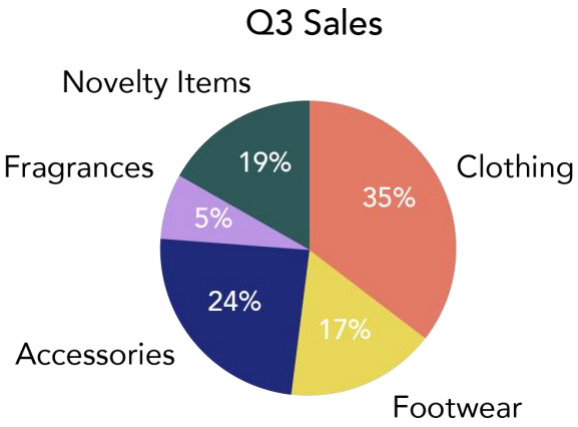
(4)

Stacked bar charts can be used to show the proportions in different categories.



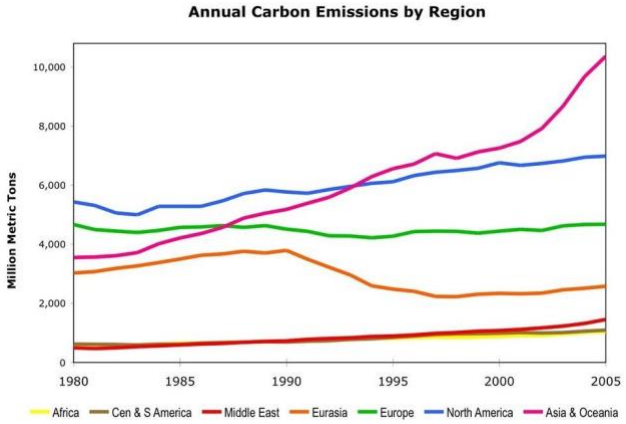
(5)

A **pie chart** should only be used for categorical data to show the proportions of your sample in different groups. It is most effective when you have from 3 to 8 groups.



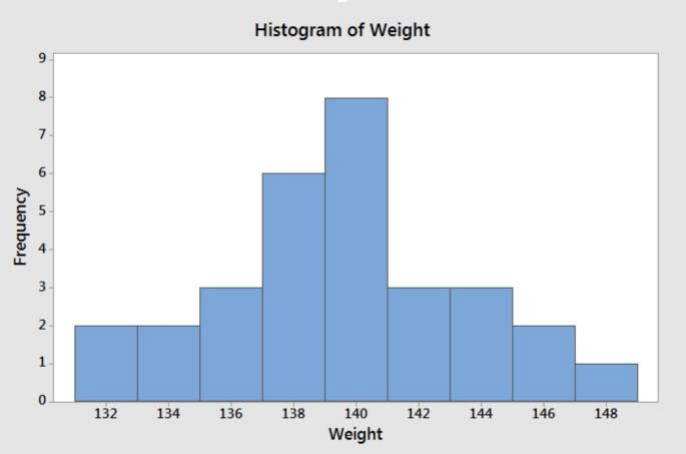
(6)

A **line graph** is most appropriate to show changes over time. You can include more than one line on the same graph to compare different measurements.



(7)

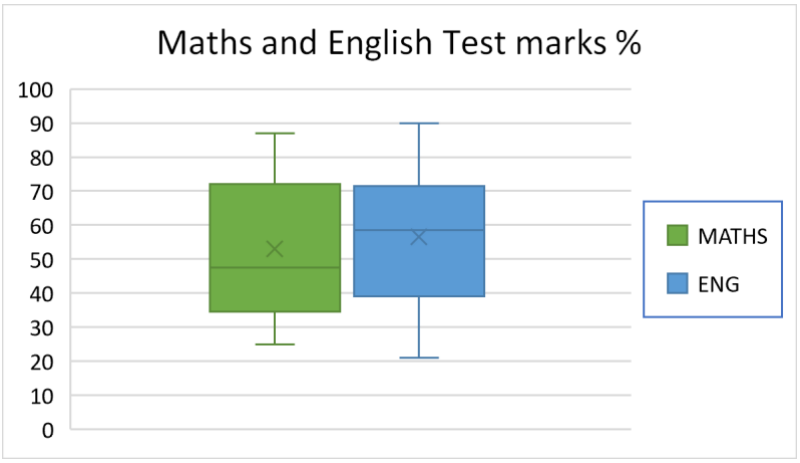
A **histogram** shows the shape of scale data and can be used to check whether the data looks normally distributed.



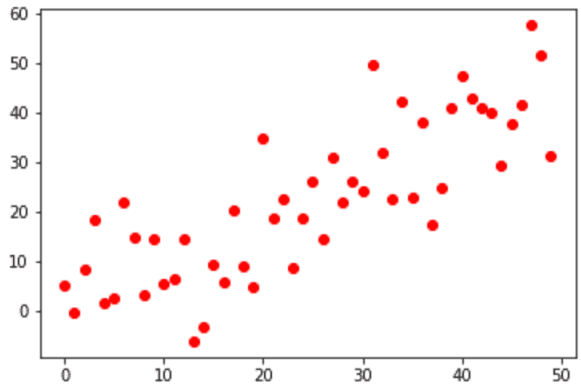
(8)

1

Box plots summarise scale data, showing the maximum and minimum measurements, the median and upper and lower quartiles. (Sometimes they also show the mean and any outliers.)



A **scatter graph** illustrates the association between two scale variables.
You might use a scatter graph to see if the association looks linear and you could add a regression line (line of best fit).



(9)

Picture credits

1. http://alphastockimages.com/ Original Author: Nick Youngson .Original Image: https://www.thebluediamondgallery.com/tablet-dictionary/a/average.html (CC BY-SA 3.0)
2. https://faculty.elgin.edu/dkernler/statistics/ch03/images/median.jpg (CC-BY-NC-SA)
3. https://flickr.com/photos/dw2002/5451587646 (CC BY-NC-ND 2.0)
4. https://www.stats4stem.org/r-barplots (CC BY-NC-SA 4.0)
5. https://www.flickr.com/photos/meadowsaffron/24494446911/ (CC BY-NC-SA 2.0)
6. https://assets.coursehero.com/study-guides/lumen/images/wmopen-businesscommunicationmgrs/charts-diagrams-and-graphic-organizers/Charts_PieCharts3.png
7. https://www.flickr.com/photos/mplemmon/3203403862/in/photostream/ (CC BY-SA 2.0)
8. https://online.stat.psu.edu/stat500/lesson/1/1.6/1.6.2 (CC BY-NC 4.0)
9. https://vrzkj25a871bpq7t1ugcgmn9-wpengine.netdna-ssl.com/wp-content/uploads/2019/01/simple-matplotlib-scatter-plot-red.png