Exploratory data analysis (EDA)
Part 5: Basic plots

By: Noureddin Sadawi, PhD

University of London

Bar chart

- A bar is used to visualise categorical data with bars.

- The categorical data needs to have corresponding values.

- The lengths (or heights) of such bars are proportional to those values.

- The bars can be plotted vertically or horizontally.

Bar chart

- They are used to display some numeric measure (e.g. number, frequency, mean) for different categories or discrete groups.

- They are flexible as they can be constructed in various ways.

  - For example, sometimes the same category can have more than one set of values and in this can the bars can be grouped.

Bar chart
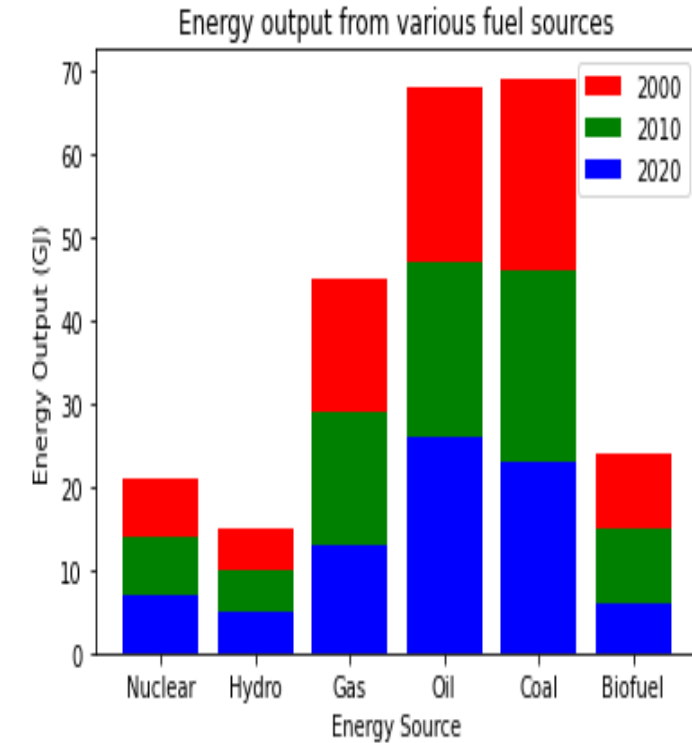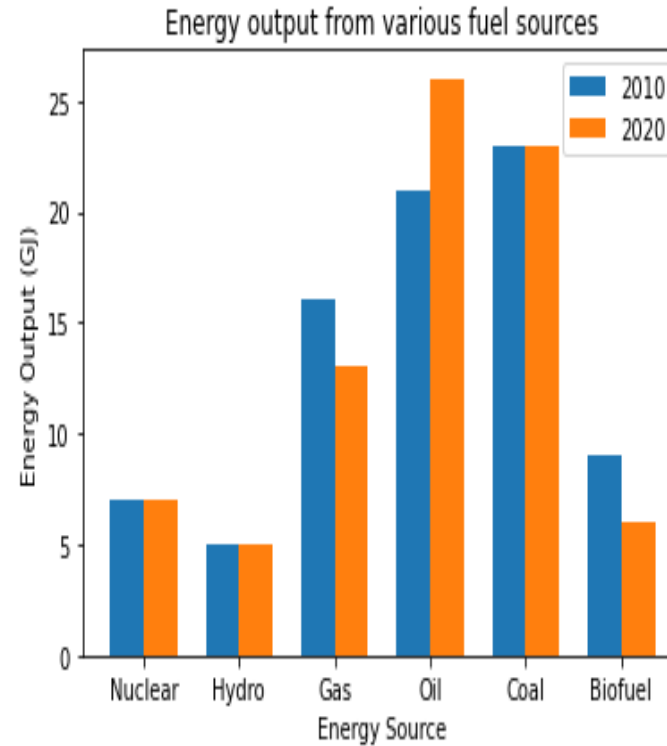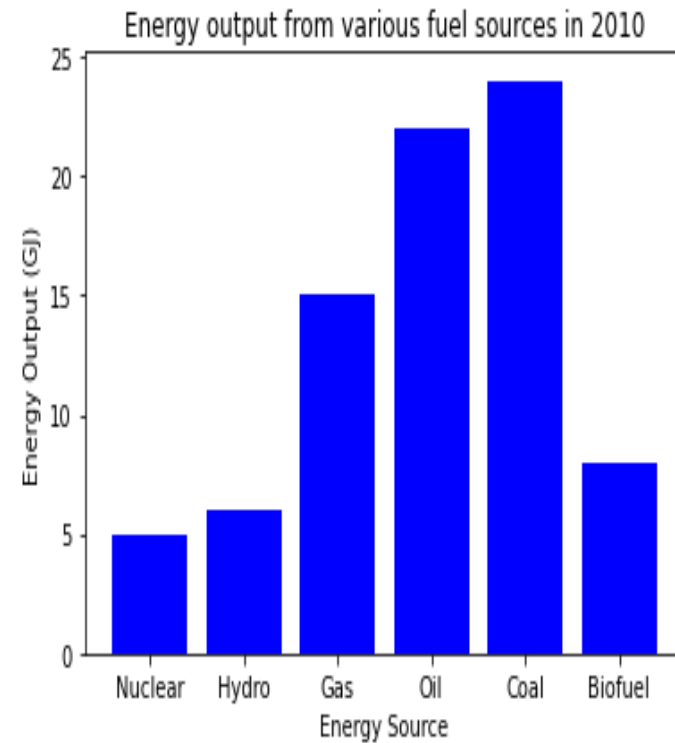
Histogram

- A histogram plot accurately displays the distribution of numerical data.

- It consists of several bars that show the frequency of a corresponding value.

- To create a histogram we divide the data into a series of non-overlapping intervals and then count how values is in each interval.

- These intervals are referred to as the **bins** of the histogram.

Histogram

- Histograms are a special form of bar charts.
- They are used to plot one continuous variable (compare with bar charts).
- For example a histogram could be used to present details of the marks of all students in a certain module (because the mark is a continuous rather than a discrete category).
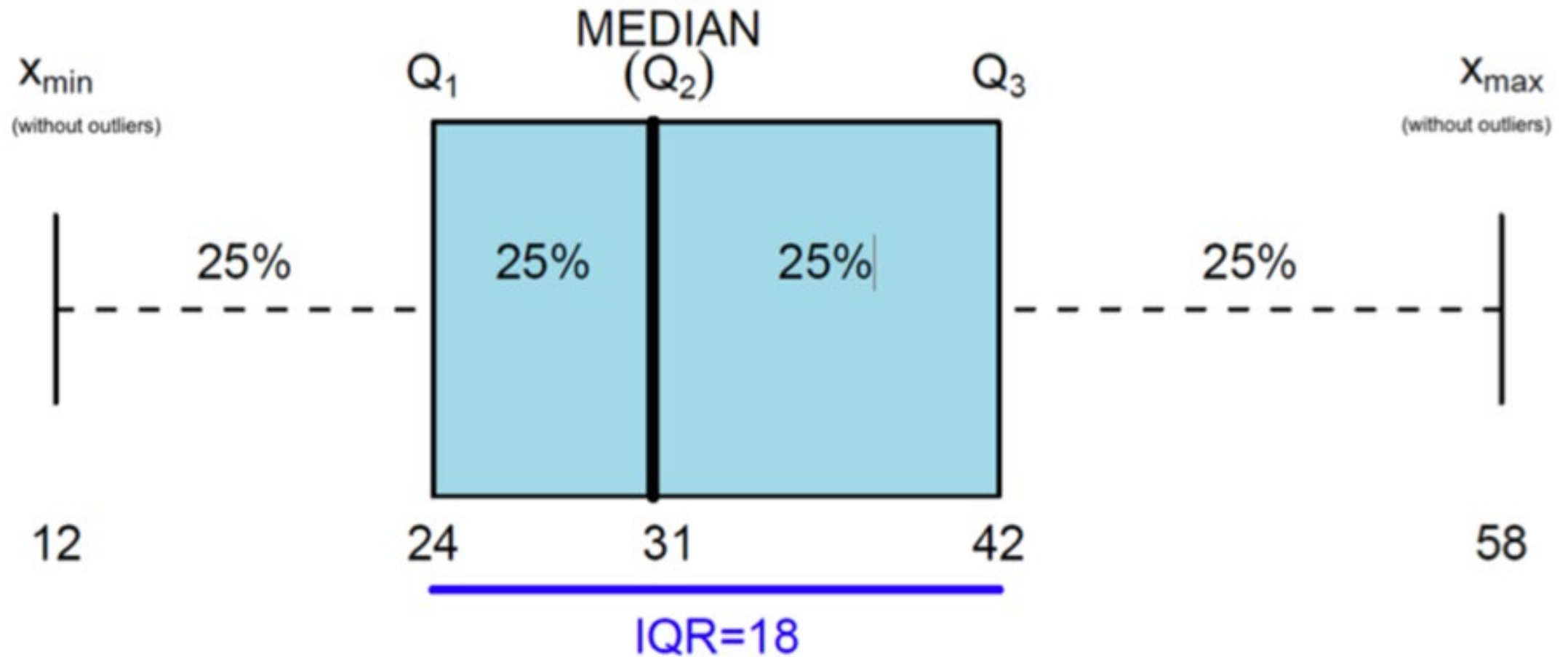
Histogram
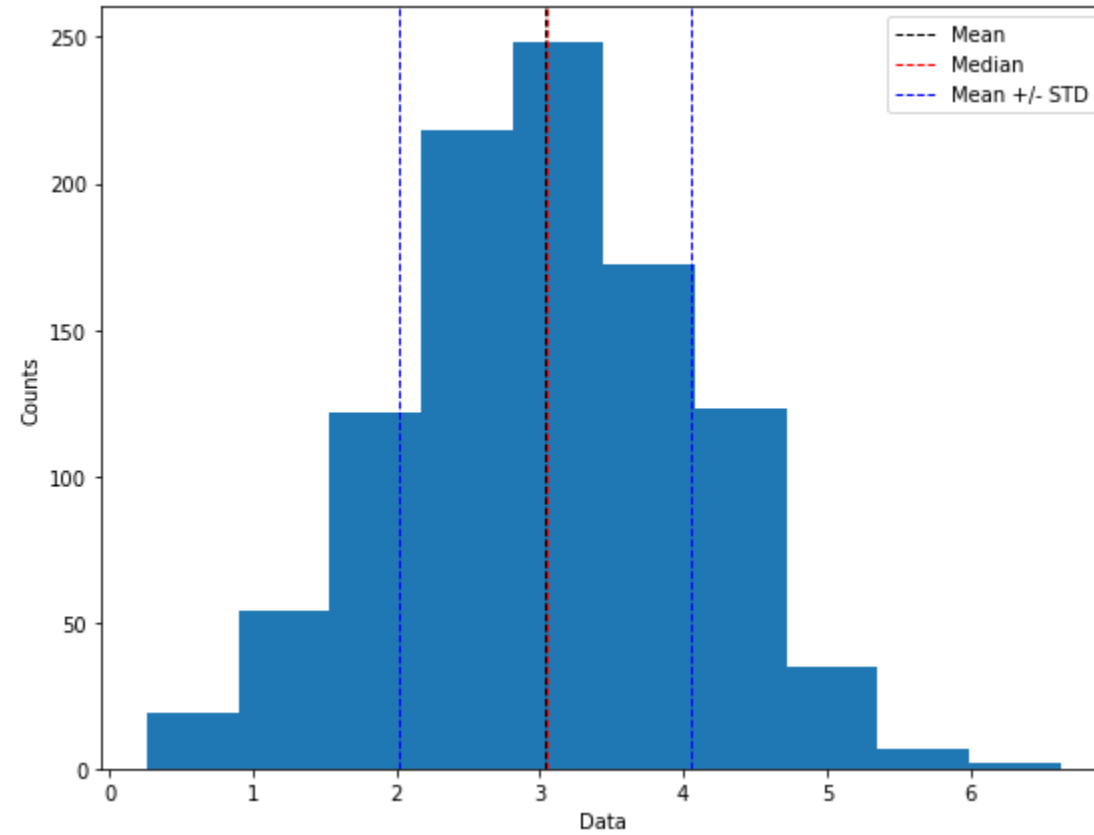
Boxplot

- A boxplot is a way of graphically displaying numerical data through its quartiles.

- It often has whiskers (lines extending from the box) that indicate variability outside Q1 and Q3 (i.e. lower and upper quartiles).

- Outliers can be depicted as individual points.

- You can also view the range and interquartile range on a boxplot.

Boxplot

- Illustrates the three quartiles and any extreme values (i.e. outliers).

- Designed to give an easy to read representation of the location and spread of a distribution.

- If there are one or more outliers, they are plotted separately as points on the chart.

Boxplot

Boxplot