Creating Box and Whisker plots in R

Ralph Mansson

Introduction

Box and Whisker plots

The box and whisker plot is a display that is used to summarise data via a graphical representation of a five number summary of a set of data. The summary statistics are the median of the data, lower and upper quartiles and the minimum and maximum values. The box and whisker plot is an effective way to study the distribution of a set of data and the shape to determine skewness and to compare distributions between groups without making any assumptions about the underlying distribution. Extreme values at either end of the scale, small or large, are sometimes included on the display.

To illustrate box and whisker plots we will consider meteorological data collected monthly at Southampton, UK between 1950 and 1999. This data is available from http://www.metoffice.gov.uk/ and we will compare the range of temperatures recorded in each month of the year over this period by creating box and whisker plots. The data is assumed to have been imported into R and stored in a data frame called soton.df.

Base Graphics

The base graphics approach uses the function boxplot to create box and whisker plots. This function can be used with a formula rather than specifying two separate vectors of data and a data frame can be included in the function to point towards a source of data to be used. The horizontal and vertical axes labels are specified with the xlab and ylab arguments respectively.
and the title of the plot is created using the main argument.

The box and whisker plot is shown here.

The boxplot makes it easy to create a reasonable box and whisker plot.

The function boxplot makes it easy to create a reasonable box and whisker plot.

The distinctive gray background differs from the default clear background used in the other two approaches. The boxes themselves have a cleaner look in this graph than the other two.

The lattice version of box and whisker plots is shown here.

The lattice version of the graph is shown here.