Learning Enhancement Team

University of East Anglia

STUDENT SUPPORT SERVICE

Steps into Statistics

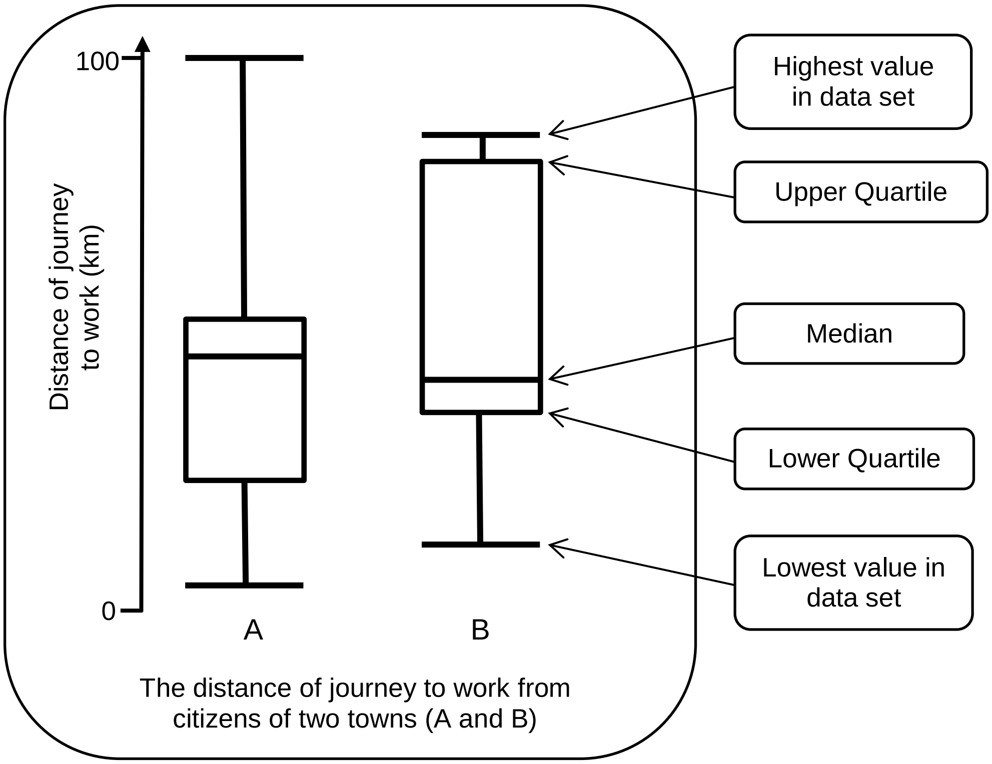
Box-and-Whisker Plots

This guide describes box-and-whisker plots and how to interpret them. It also discusses the levels of data that are appropriate to use them with and situations in which to use them.

Introduction

There are many different ways to present data in statistics, the most common are bar charts, pie charts, and histograms (see study guides: *Bar Charts*, *Pie Charts* and *Histograms*). One fairly recent addition to these diagrams is the **box-and-whisker plot**, sometimes called simply a **boxplot**. They were devised by John Tukey in 1977 and offer a simple and effective way of presenting and comparing data sets which do not follow an underlying distribution (**non-parametric data**). Often box-and-whisker plots are used to present **ordinal** (ranked) data, see study guide: *Levels of Data*.

A typical box-and-whisker plot is shown below and is annotated to show you what each of the parts represent. You can clearly see why it is called a box-and-whisker plot.



The distance of journey to work from citizens of two towns (A and B)

Features of box-and-whisker plots

There is no convention to say that box-and-whisker plots must be vertical, like the one on the previous page of this guide, and you will often see horizontal plots. However, **for vertical plots**:

- The **top whisker** is the **highest value** in the data set.*
- The **bottom whisker** is the **lowest value** in the data set.*
- The **top of the box** is the **upper quartile** of the data set.
- The **bottom of the box** is the **lower quartile** of the data set.
- The **line in the box** is the **median** of the data set.

* The positions of the upper and lower whiskers can be defined in different ways and these are discussed in the final section of this guide.

These definitions are easily applied to horizontal box-and-whisker plots. The study guide: *Measurements of Central Tendency* has more information about the median. You can find out more about the upper and lower quartiles in the study guide: *Measurements of Spread I: Range and Interquartile Range*.

One of the great strengths of box-and-whisker plots is that they show a great deal of the descriptive statistics of a data set in a plain and clear way. For example:

- The **difference between the top whisker and the bottom whisker** is the **range** of the data set.
- The **height of the box** is the **inter-quartile range**.
- The **height of the box** represents the middle **50% of the data**.
- The **whisker above the box** represents the top **25% of the data**.
- The **whisker below the box** represents the bottom **25% of the data**.
- If the **box is closer to the bottom whisker** then, in general, the data are **positively skewed**.**
- If the **box is closer to the top whisker** then, in general the data are **negatively skewed**.**

You can find out more about the range and inter-quartile range in the study guide: *Measurements of Spread I: Range and Interquartile Range*.

** **Skewness** is the statistical measure of how much a data set bunches to towards higher or lower values. Its definition is beyond the remit of this guide but you can talk to a **Learning Enhancement Tutor** if you want to understand more about skewness.

Using a box-and-whisker plot to compare data sets

Box-and-whisker plots are an excellent visual aid to help you to compare data sets. For example, the plot on the first page of this guide can be used to make the following statements:

- The spread of journeys to work in town A is greater than in town B. (The Range for town A is greater than for town B.)
- The median journey to work is slightly shorter from town B than from town A.
- The middle 50% of journeys encompass a narrower spread of distances in town A than town B. (The Interquartile Range is smaller for town A.)
- There is evidence that the data for town A are positively skewed.
- There is evidence that the data for town B are negatively skewed.
- The top 25% of journeys have a much larger spread for town A compared to town B.
- The bottom 25% of journeys for each town have roughly the same spread.

Although these statements are not the result of statistical testing, having a good grasp of how data sets compare to each other in this way helps you understand any underlying trends in more depth.

Adaptations of box-and-whisker plots

Even though the upper and lower quartiles are always used as the boundaries for the box, the position of the whiskers can vary. The plot on the first page of this guide uses the highest and lowest data values as the whiskers. There are other ways the whiskers can be positioned such as:

- One standard deviation above and below the mean.
- The lowest data point 1.5 times the inter-quartile range below the lower quartile and the highest data point 1.5 times the inter-quartile range above the upper quartile.
- The x^{th} and $(100-x)^{\text{th}}$ percentiles (such as 2nd and 98th or 5th and 95th).

Other information is often included such as an “X” to denote the position of the mean (where appropriate).

Both outliers and data which is not included directly in the analysis can be accommodated by box-and-whisker plots. They are often shown outside of the whiskers and denoted by either a circle ° or an asterisk *. If you are excluding data from your plot you must ensure that you write why you are doing so in the text which accompanies the plot. You should also give a key which explains any extra symbols that you use.

Want to know more?

If you have any further questions about this topic you can make an appointment to see a **Learning Enhancement Tutor** in the **Student Support Service**, as well as speaking to your lecturer or adviser.

☎	Call:	01603 592761
💻	Ask:	ask.let@uea.ac.uk
🖱	Click:	https://portal.uea.ac.uk/student-support-service/learning-enhancement

There are many other resources to help you with your studies on our website.
For this topic there is a webcast.

Your comments or suggestions about our resources are very welcome.

Scan the QR-code with a smartphone app for a webcast of this study guide.