



R Girls School



---

title: "Boxplots"

author: "Put your name here"

date: "`r format(Sys.time(), '%d %B, %Y')`"

output: html\_document

---

``{r setup, include=FALSE}

knitr::opts\_chunk\$set(echo = TRUE)

library(tidyverse)

library(astsa)

#<https://donsteward.blogspot.com/search/label/boxplots>

---

### GCSE KS4: Representing and analysing data using box plots

### Lesson objectives

To understand how to visualise the shape of data using box plots\

To produce box plots and the related five number summary\

To interpret box plots\

### Success criteria

\* Draw a box plot using R code

\* Produce a five number summary using R code

\* Read data from a box plot

\* Compare two box plots

### ### Keywords

box plot,

minimum, maximum, lower quartile, median, upper quartile, five number summary, inter-quartile range, range

### ### Remember

Knit your R markdown document frequently. Knit for the final time when you have completed the lesson. This final output file will be the one your teacher marks.

Write your code in the code chunks.

You can run your R code chunk in the Markdown file by clicking on the little arrow on the right of the chunk.

### #### What is a box plot

A box plot is a way to show the shape or distribution of a set of data.

It shows five useful features of the data, known as the five number summary.

Minimum - the smallest value\

Maximum - the largest value\

Median - the middle or 50% value\

Lower quartile - the value half way between the minimum and the median or 25% value\

Upper quartile - the value half way between the median and the maximum or 75% value\

The difference between the upper and lower quartile is known as the inter-quartile range (IQR)

### #### Worked Example 1

Zaynab keeps a record of her journey times to school each morning. The times are to the nearest minute:

29,21,16,25,21,19,18,30,21,21,12,26,19,21,20,19,30,29,16,21,18,18,27,18,20

Let us draw a box plot and obtain the five number summary and the inter-quartile range (IQR). We will do this in three steps.

The R code in chunk1 tells R to store all the times into a variable called y and then to use these times to draw a box plot of the data.

```

```{r chunk1}

y <- c(29,21,16,25,21,19,18,30,21,21,12,26,19,21,20,19,30,29,16,21,18,18,27,18,20)

boxplot(y, xlab="Times (mins)", horizontal=T)

```

```

To help us read values from the box plot it is useful to have gridlines on the plot.\

Here is a simple trick to get gridlines on the plot.

The code in chunk2 tells R to use the numbers it store in variable y to draw a box plot and then to add grid lines.

Click on the little arrow on the right of the code chunk. The box plot will be drawn below the chunk.

```

```{r chunk2}

# we can calculate the five number summary like this

boxplot(y, xlab="Times (mins)", horizontal=T)

Grid()

boxplot(y, xlab="Times (mins)", horizontal=T, add=T)

```

```

We can also get R to calculate the five number summary for us.

```

```{r chunk3}

# we can calculate the five number summary like this

summary(y) # five number summary but also provides the mean

IQR(y) # inter-quartile range

```

```

#### #### Activity 1:

Produce a box plot, the five number summary and IQR of the following data

Aisha records the length of time spent by each member of her class on their last English homework assignment.

The times are recorded to the nearest minute.

85,124,55,140,120,61,95,105,118,180,55,78,130,112,70,126,60,90,115,60,142,100,105,65,100,75

Use the R code in chunk1 to help you draw the box plot. Remember to replace the example data with the new dataset.

```
``{r chunk4}
```

```
...
```

Now use the code in chunk2 to add grid lines to your box plot

```
``{r chunk5}
```

```
...
```

Now calculate the five number summary using code in chunk3

```
``{r chunk6}
```

```
...
```

#### #### Worked Example 2

Below are the percentage exam marks out of 100 for Maths and English.

Maths (%): 98,79,51,54,62,61,56,87,70,60,93,51,52,54,68

English (%): 37,50,58,45,93,47,47,45,38,61,65,46,97,99,54

Produce two box plots side by side

```
```{r chunk7}
```

```
maths <- c(98,79,51,54,62,61,56,87,70,60,93,51,52,54,68)
```

```
english <- c(37,50,58,45,93,47,47,45,38,61,65,46,97,99,54)
```

```
boxplot(maths, english, xlab="Marks%", names=c("Maths", "English"), horizontal=T)
```

```
Grid()
```

```
boxplot(maths, english, xlab="Marks%", names=c("Maths", "English"), horizontal=T, add=T)
```

```
summary(maths)
```

```
summary(english)
```

```
IQR(maths)
```

```
IQR(english)
```

```
```
```

Write down the five number summaries below

What are the inter-quartile ranges

#### Activity 2:

Below are the heights of boys and girls of a similar age.

Boys (cm): 181,157,159,179,186,159,178,162,137,184,140,173,176

Girls (cm): 172,151,176,159,139,179,178,162,134,166,164,172,170

Produce two box plots side by side with a grid and answer the following questions.

Remember to replace the example data with the new information (boys, girls, heights, data)

```
```{r chunk8 Activity2}
```

```
```
```

## Questions

Are the statements below True or False - and explain why

1. The girls are taller on average
2. Half the girls are over 165 cm tall
3. The girls show less spread in height
4. The boys show less spread in height
5. The shortest person is a girl
6. The tallest person is a boy
7. Half the boys are over 172 cm tall

8. Half the girls are under 165cm tall

KNIT YOUR DOCUMENT for the final time. This will be the version your teacher will mark

#### The END