Comparing Box and Whisker Plots

When comparing box plots it is important to look for how many comparisons they want.
The number of marks on offer tells you how many comparisons they want. i.e. 3 marks means 3 comparisons
You need to do your best to comment on different aspects of the box plots rather than all on the same thing. You must also make sure that you use the correct statistical language

The four main comparison areas are '*SASKIa*':

*S*pread	- this refers to the ranges or interquartile ranges calculated from the box plots.
*A*verages	- this refers to the medians seen on the box plots
*Sk*ew	- this refers to the skews* seen on the box plots
*I*nterpretation	- this refers to an explanation of the spread or average *IN CONTEXT*
*a*nomalies	- this refers to an outliers

*See skew for separate notes on this

*Sp*read

Use the scale to work out the values for the range and IQR in each box plot.
State which box plot has the largest spread
Decide which is the most significant (usually you will see a more significant difference between the box plots in one than the other)

*A*verages

Use the scale to work out the value for the median in each box plot
State which box plot has the highest median

*Sk*ew

Positive Skew

The median is closer to the lower quartile than the upper quartile
This means that although the range of the data is wide, the average of the data is towards the lower end of the scale

Negative Skew

The median is closer to the upper quartile than the lower quartile
This means that although the range of the data is wide, the average of the data is towards the upper end of the scale

Neither of these distributions are better than the other, it just helps us describe the distribution of the data and compare them

*I*nterpretation

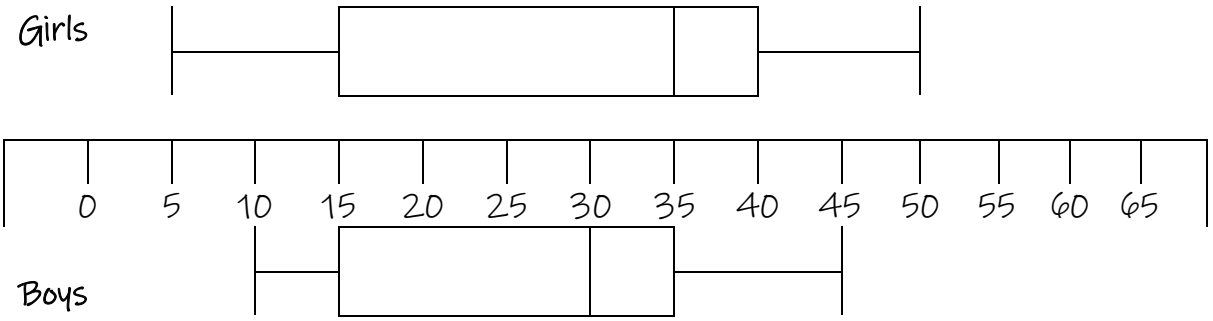
Comment on how much more '*spread out*' the data is for one set or how the data is spread similarly for both sets
Comment on which set of data is higher/lower/faster/slower etc '*on average*'

*a*nomalies

You may also wish to mention any outliers that can be seen with an 'x' on the box plots

For each of the following pairs of box plots, use the SASkIa acronym to write a comparison of the two data sets.

A box pot showing the summarized data for test scored of girls and boys out of a possible 65 marks



S: _____

_____

A: _____

_____

Sk: _____

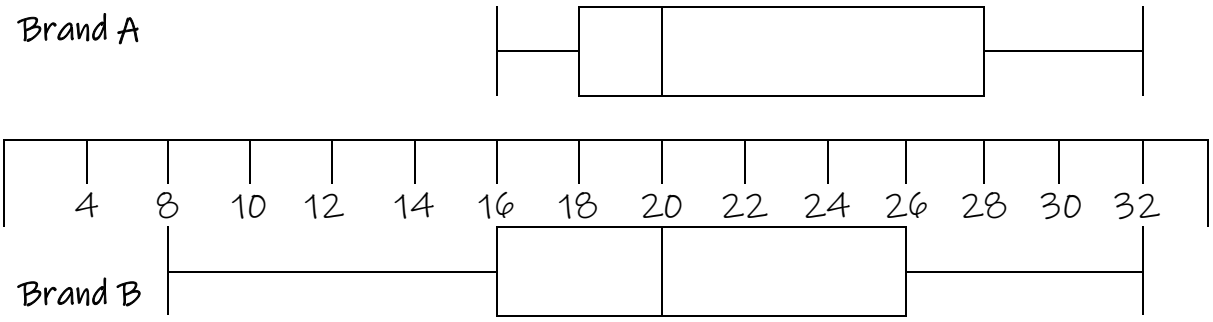
_____

I: _____

_____

a: _____

A box pot showing the summarized data for weights (g) of mice who are using two different brands of food.



S: _____

_____

A: _____

_____

Sk: _____

_____

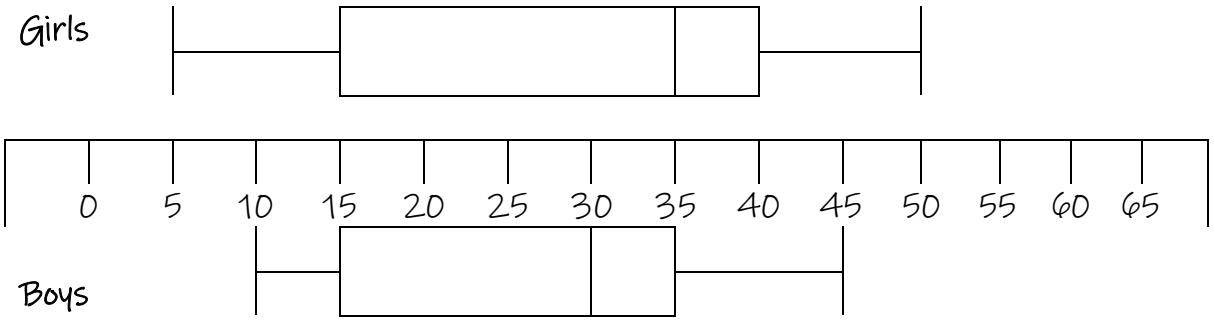
I: _____

_____

a: _____

Solutions

A box pot showing the summarized data for test scored of girls and boys out of a possible 65 marks



S: Girls have both a larger **RANGE** and **INTERQUARTILE RANGE** than the boys

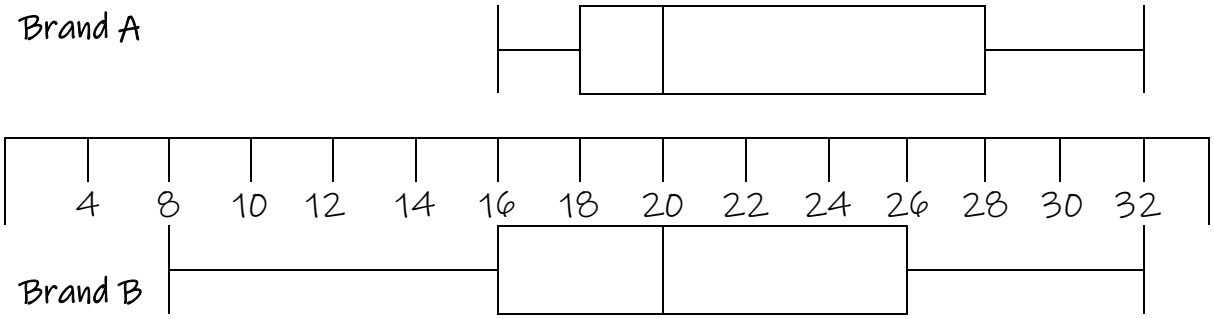
A: Girls have a higher **MEDIAN** than the boys

Sk: Bothe the girls and the boys show a **NEGATIVE SKEW**

I: The girls test scores were more **SPREAD OUT** than the boys test scores **ON AVERAGE** the girls scored higher on the test than the boys

a: No outliers

A box pot showing the summarized data for weights (g) of mice who are using two different brands of food.



S: Brand B has a larger **RANGE** than Brand A
Both Brand A and Brand B have the same **INTERQUARTILE RANGE**

A: Both Brand A and Brand B have the same **MEDIAN**

Sk: Both Brand A and Brand B show a **POSITIVE SKEW**

I: The weights of the mice who are eating Brand B food are more **SPREAD OUT ON AVERAGE** the weights of the mice who eat Brand A are the same as those who eat Brand B

a: No outliers