Statistical measures/tools that can be used to describe/analyze data sets

- Compare averages and minimum and maximum values
- Calculate and comment on range
- Comment on mode (most common value) if you can
- Identify and comment on outliers and clusters
- Compare shapes and comment on distribution - Symmetrical or skewed
- Draw percentage comparisons from double box and whisker plots

Indicators
1. A higher mean and median suggests an overall higher set of results.
2. A smaller range suggests a more consistent set of values whereas a wider range and interquartile range points to a greater spread.
3. Compare medians, if any of the pointers (minimum, maximum, lower quartile, upper quartile, median) coincide, what do they mean in comparison with each other etc
4. Refer to tasks on MHOL week 6 and 7 **(Answers for interpreting double box and whisker plots - ideas for analysis of data)**

Statistical Investigation for Practice Data (Male and Female school bag weights)

Problem/Investigation question - I wonder if the school bag weights of females tends to be higher than the males OR I am investigating the school bag weights of females in comparison to males OR Do females carry more bag weight than males in general?

Plan - I will be comparing the bag weights (in kgs) of males and females through comparative analysis of 2 different graphs.
Data - The data source is USA Statistics.
Analysis and Conclusion - I will be writing a summary of findings in response to my question based on an analysis of the graphs.

Description and Analysis

The average bag weight of females is 8kgs which is double that of the males.(4kgs). The higher mean and median suggests an overall higher bag weight for females.

The maximum bag weight of females is 24kgs whereas the maximum bag weight of males is 7kgs. With this comparison I can see there is a difference of 17kgs. In saying that, as can be seen from the graph the 24kg bag weight is certainly looking like an outlier and so is 22kgs. Hence, the difference in maximum weights is unlikely to be a true representation of the data set. If we exclude the outliers and take 12kgs to be the maximum weight of bags for the females then the difference is 5kgs which is more plausible.

The most common bag weight for females is 8kgs and for males it is 4kgs which also coincides with their medians and the graph displays a cluster around those weights.

The female box plot has a longer whisker on the right which shifts the graph further up the scale but this can be attributed to the outliers. Otherwise, both graphs display, more or less, a symmetrical distribution. A smaller range of 6kgs (max value - min value) for the males shows a more consistent set of bag weights as compared to the females which shows more variation.

I notice from the graphs that there is no overlap between the boxes and the median bag weight of the females lies outside the box of males.The lower quartile of the females coincides with the maximum value of the males. It means that 75% of the females have a higher bag weight than all (100%) of the males.

Conclusion - From my comparative analysis of the graphs of males and females bag weights I can conclude that the females tend to have a higher bag weight than the males in general.

Interesting feature and informal inference - The bag weight for two females being 22kgs and 24kgs is quite unusual in terms of students to carry such a heavy bag to school. The data would require further investigation regarding this aspect. Information about the age group of the students and what a typical school day looks like for them would be helpful for a more comprehensive analysis. It could have been a recording error or these students might be carrying extra weight for a special/outdoor activity on that particular day.