# Plotting For Success

## Teacher Notes

VICTORIAN BIOSCIENCE EDUCATION CENTRE

## Curriculum Links

**Level 10 Mathematics**

- Determine quartiles and interquartile range and investigate the effect of individual data values, including outliers on the interquartile range (VCMSP349).
- Construct and interpret box plots and use them to compare data sets (VCMSP350).

**General Mathematics Units 1 and 2 Area of Study 6: Statistics**

Investigating and comparing data distributions. This topic includes:

- Measures of centre and spread and their use in summarising numerical data distributions, including use of and calculation of the sample summary statistics, median, mean, range, interquartile range (IQR) and standard deviation; and choosing between the measures of centre and spread.
- The five-number summary and the boxplot as its graphical representation and display, including the use of the lower fence (Q1– 1.5 × IQR) and upper fence (Q3 + 1.5 × IQR) to identify possible outliers.
- Use of back-to-back stem plots or parallel boxplots, as appropriate, to compare the distributions of a single numerical variable across two or more groups in terms of centre (median) and spread (IQR and range), and the interpretation of any differences observed in the context of the data.

## Learning Intention and Success Criteria

**Learning intention**

To understand how to use boxplots and summary statistics to compare two data distributions.

**Success criteria**

- ✓ Approximate five point summary statistics from a boxplot.
- ✓ Calculate range.
- ✓ Describe the shape of a data distribution.
- ✓ Interpret a boxplot the answer questions about data sets.
- ✓ Explain when and why median is a more appropriate measure of spread than mean.

© 2020 BioLAB: The Victorian BioScience Education Centre

# Plotting For Success

## Instructions to Teachers

Athletes, coaches, engineers and sports scientists rely on the collection and analyses of data in order to develop new technologies that will improve equipment and performance. Data analysis skills are integral to the success of this process. This data analysis task was developed for students to apply their data analyses skills in the context of athlete training.

The task would be suitable for use at Level 10 or VCE General Mathematics. The questions are accessible to both levels; however, depth of expected answers at each level would differ.

## Solutions

**Describing Data:**

1. Answers will differ slightly.

| Statistic | 2019 | 2020 |
|---|---|---|
| Minimum Value (Min X) | 19: 05 | 18: 35 |
| Lower Quartile (Q1) | 20: 00 | 19: 20 |
| Median (Q2 ) | 20: 50 | 20: 30 |
| Upper Quartile (Q3) | 22: 05 | 22:15 |
| Max X | 22: 50 | 24: 00 |
| Range (Max X – Min X) | 2: 45 | 5: 25 |

2. The best training run was in 2020, since the lowest time was about 18 min 35 seconds whereas the best time in 2019 was over 19 minutes.
3. 2019 had the most consistent times because the range is less than 2020 and so is the interquartile range. This means that not only was the difference between the best and worst times less than 2020, but 50% of the runs were closer in time to each other.
4. The injury probably occurred during 2020. The right-hand whisker for this plot indicates that some of the runs were quite slow, while the short left-hand whisker indicates that the faster runs were quite consistent in times and quite fast.
5. The distribution for 2019 is fairly symmetrical with no long tail or stretching of the data on either side of the median. The distribution for 2020 is positively skewed (probably due to an injury). There is an elongated tail of data to the right of the median.

**Challenge:**

There are a number of possible answers. Data should have a minimum, maximum and median value that matches the table and will need to draw on their understanding of quartiles to complete the data set.

© 2020 BioLAB: The Victorian BioScience Education Centre