Comparing Data Sets

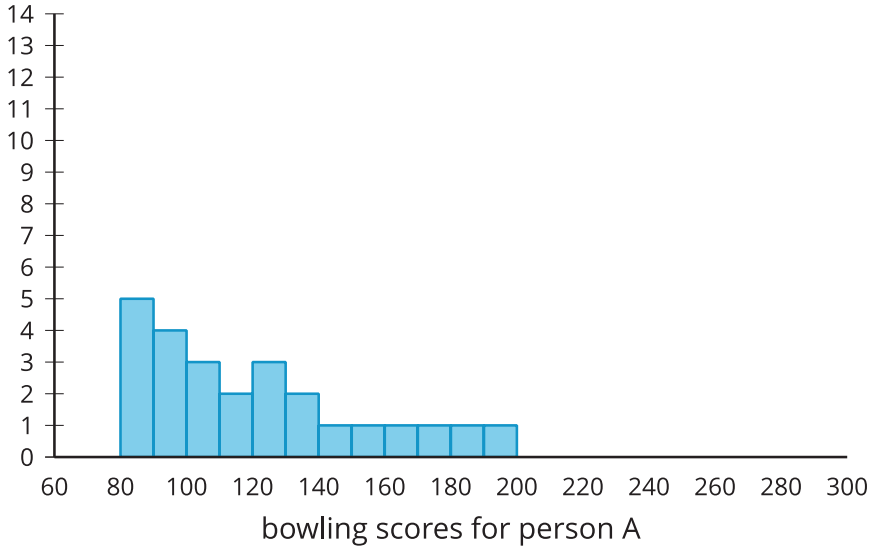
Let's compare statistics for data sets.

15.1 Bowling Partners

Each histogram shows the bowling scores for the last 25 games played by each person. Choose 2 of these people to join your bowling team. Explain your reasoning.

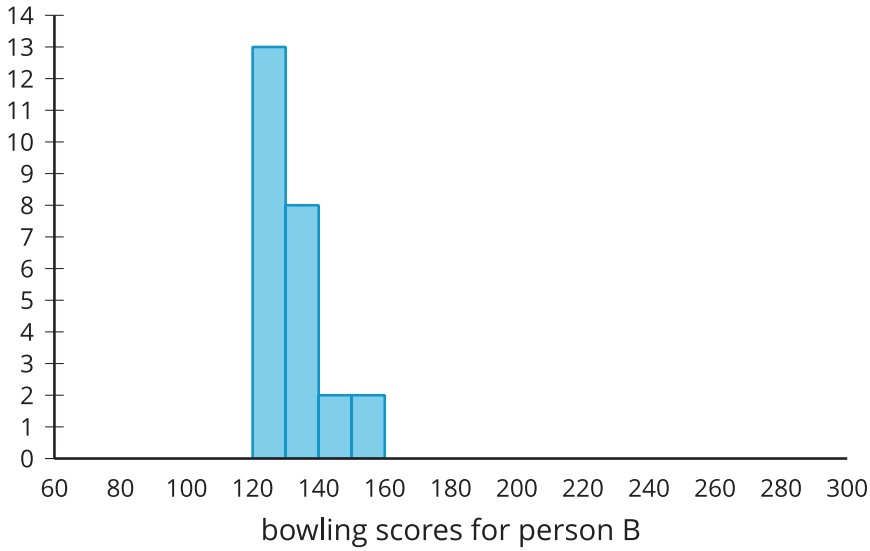
Person A

- mean: 118.96
- median: 111
- standard deviation: 32.96
- interquartile range: 44



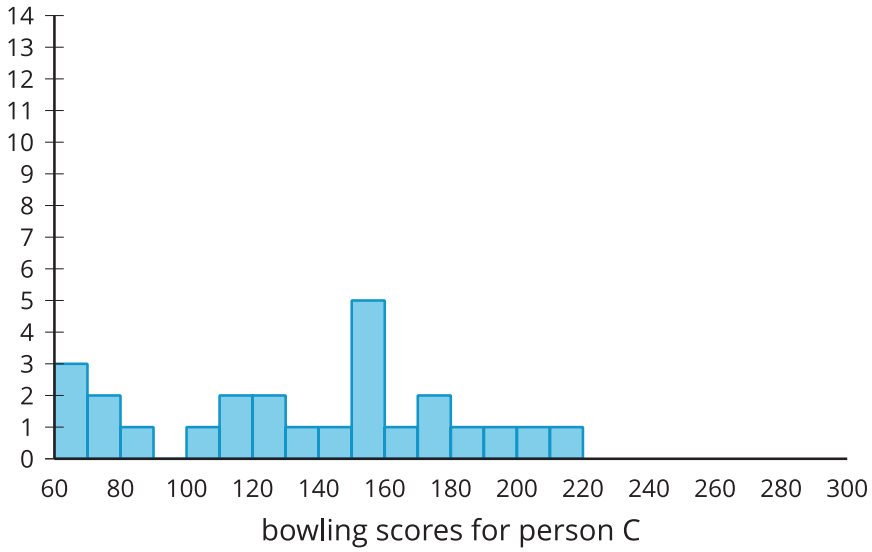
Person B

- mean: 131.08
- median: 129
- standard deviation: 8.64
- interquartile range: 8



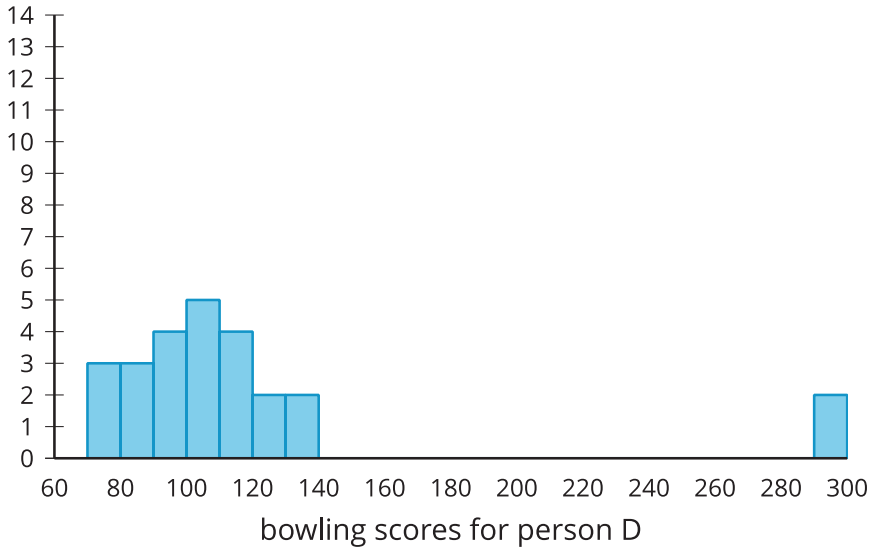
Person C

- mean: 133.92
- median: 145
- standard deviation: 45.04
- interquartile range: 74



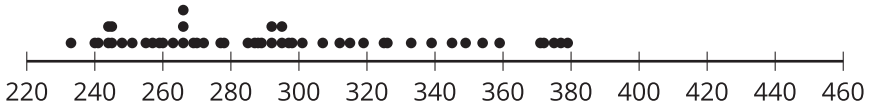
Person D

- mean: 116.56
- median: 103
- standard deviation: 56.22
- interquartile range: 31.5



15.2 Comparing Marathon Times

All of the marathon runners from each of two different age groups have their finishing times represented in the dot plot.



ages 30 – 39 marathon finish times (minutes)



ages 40 – 49 marathon finish times (minutes)

1. Which age group tends to take longer to run the marathon? Explain your reasoning.

2. Which age group has more variable finish times? Explain your reasoning.

Are you ready for more?

1. How do you think finish times for a 20–29 age range will compare to these two distributions?

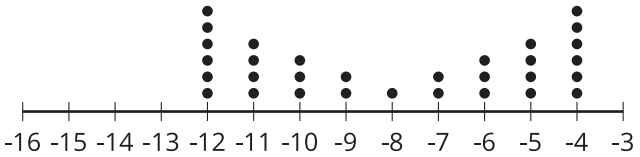
2. Find some actual marathon finish times for this group and make a dot or box plot of your data to help compare.

15.3 Comparing Measures

For each group of data sets,

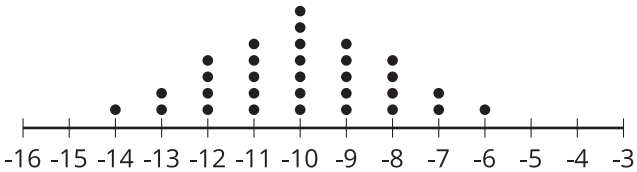
- Determine the best measure of center and measure of variability to use based on the shape of each distribution.
- Determine which set has the greatest measure of center.
- Determine which set has the greatest measure of variability.
- Be prepared to explain your reasoning.

1a



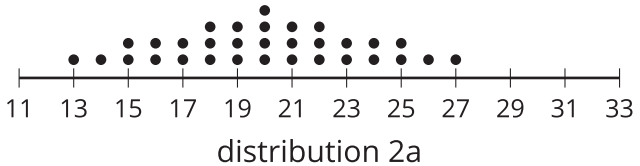
distribution 1a

1b



distribution 1b

2a



distribution 2a

2b



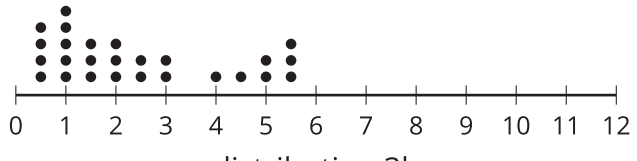
distribution 2b

3a



distribution 3a

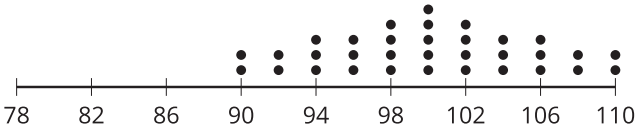
3b



distribution 3b

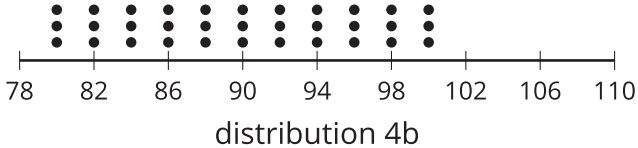
CC BY NC 2024
Illustrative Mathematics®

4a



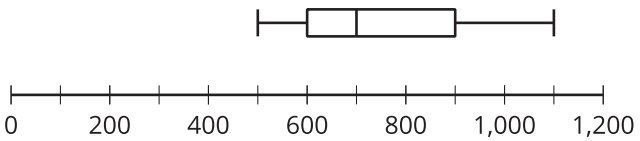
distribution 4a

4b



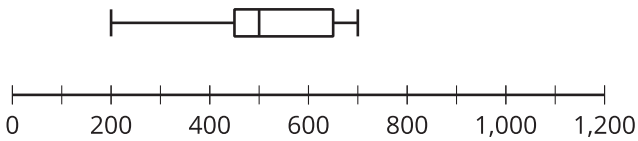
distribution 4b

5a



distribution 5a

5b



distribution 5b

6a

A political podcast has mostly reviews that either love the podcast or hate it.

6b

A cooking podcast has reviews that neither hate nor love the podcast.

7a

Stress testing concrete from site A has all 12 samples break at 450 pounds per square inch (psi).

7b

Stress testing concrete from site B has samples break every 10 psi starting at 450 psi until the last core is broken at 560 psi.

7c

Stress testing concrete from site C has 6 samples break at 430 psi and the other 6 break at 460 psi.

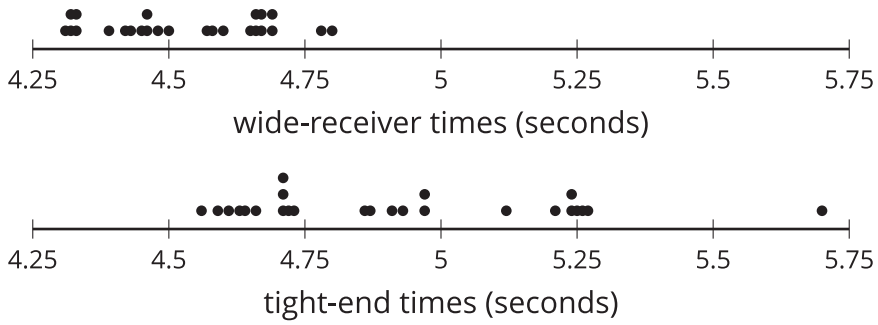
Lesson 15 Summary

To compare data sets, it is helpful to look at the measures of center and measures of variability. The shape of the distribution can help choose the most useful measure of center and measure of variability.

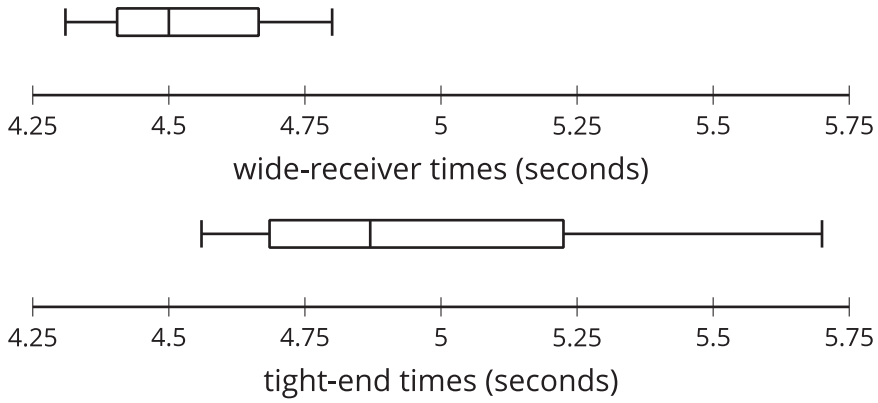
When distributions are symmetric or approximately symmetric, the mean is the preferred measure of center and should be paired with the standard deviation as the preferred measure of variability. When distributions are skewed or when outliers are present, the median is usually a better measure of center and should be paired with the interquartile range (IQR) as the preferred measure of variability.

Once the appropriate measure of center and measure of variability are selected, these measures can be compared for data sets with similar shapes.

For example, let's compare the number of seconds it takes football players to complete a 40-yard dash at two different positions. First, we can look at a dot plot of the data to see that the tight-end times do not seem distributed symmetrically, so we should probably find the median and IQR for both sets of data to compare information.



The median and IQR could be computed from the values, but can also be determined from a box plot.



This shows that the tight-end times have a greater median (about 4.9 seconds) compared to the median of wide-receiver times (about 4.5 seconds). The IQR is also greater for the tight-end times (about 0.5 seconds) compared to the IQR for the wide-receiver times (about 0.25 seconds).

This means that the tight ends tend to be slower in the 40-yard dash when compared to the wide receivers. The tight ends also have greater variability in their times. Together, this can be taken to mean that, in general, a typical wide receiver is faster than a typical tight end is, and the wide receivers tend to have more similar times to one another than the tight ends do to one another.