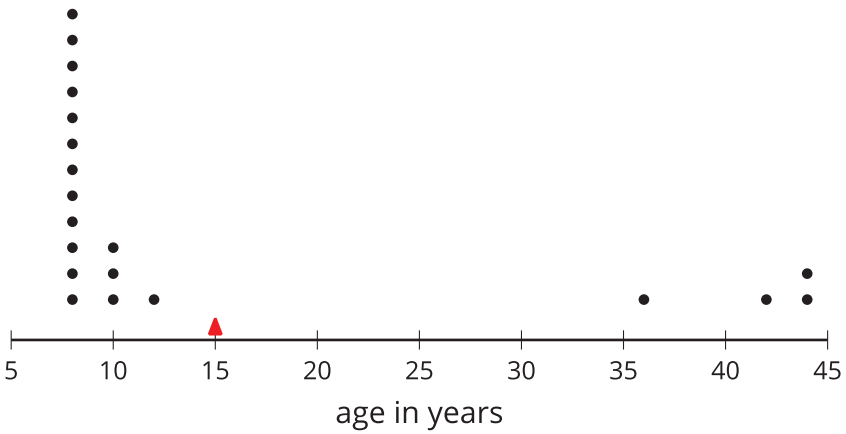
Quartiles and Interquartile Range

Let's look at other measures for describing distributions.

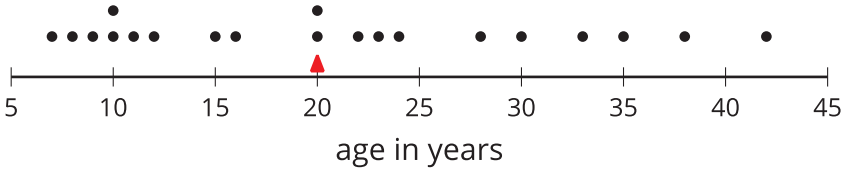
15.1 Notice and Wonder: Two Parties

Here are dot plots that show the ages of people at two different parties. The mean of each distribution is marked with a triangle.

Data set A



Data set B



What do you notice and what do you wonder about the distributions in the two dot plots?

15.2 The Five-Number Summary

Here are the ages of the people at one party, listed from least to greatest.

7 8 9 10 10 11 12 15 16 20 20 22 23 24 28 30 33 35 38 42

1. Split the data into 4 equal parts using 3 values called **quartiles**.
   a. Find the median of the data set and label it Q2 for second quartile. This splits the data into an upper half and a lower half.
   b. Find the middle value of the lower half of the data, without including the median. Label this value Q1 for first quartile
   c. Find the middle value of the upper half of the data, without including the median. Label this value Q3 for third quartile.

2. Label the least value in the set "minimum" and the greatest value "maximum."

3. The values you have identified make up the *five-number summary* for the data set. Record them here.

   minimum: ____ Q1: ____ Q2: ____ Q3: ____ maximum: ____

4. The median of this data set is 20. This tells us that half of the people at the party were 20 years old or younger, and the other half were 20 or older. What do each of these other values tell us about the ages of the people at the party?

   a. The third quartile

   b. The minimum

   c. The maximum

Are you ready for more?

There was another party where 21 people attended. Here is the five-number summary of their ages.

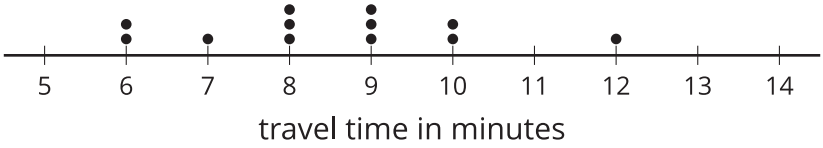
minimum: 5 Q1: 6 Q2: 27 Q3: 32 maximum: 60

1. Do you think this party had more children or fewer children than the earlier one? Explain your reasoning.

2. Were there more children or adults at this party? Explain your reasoning.

15.3 Range and Interquartile Range

1. Here is a dot plot that shows the lengths of Elena's bus rides to school, over 12 days.



   Write the five-number summary for this data set. Show your reasoning.

2. The **range** is one way to describe the spread of values in a data set. It is the difference between the maximum and minimum. What is the range of Elena's travel times?

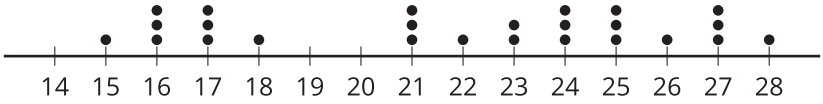
3. Another way to describe the spread of values in a data set is the **interquartile range (IQR)**. It is the difference between the third quartile (Q3) and the first quartile (Q1).

   a. What is the interquartile range (IQR) of Elena's travel times?

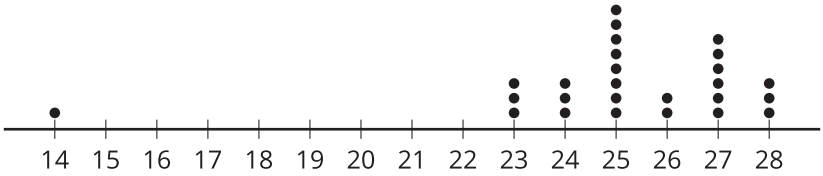
   b. What fraction of the data values are between Q1 and Q3?

4. Here are 2 more dot plots.

Data set A



Data set B



Without doing any calculations, predict:

a. Which data set has the smaller range?

b. Which data set has the smaller IQR?

5. Check your predictions by calculating the range and IQR for the data in each dot plot.

Lesson 15 Summary

Earlier we learned that the mean is a measure of the center of a distribution and the MAD is a measure of the variability (or spread) that goes with the mean. There is also a measure of spread that goes with the median. It is called the *interquartile range* (IQR).

Finding the IQR involves splitting a data set into fourths. Each of the three values that splits the data into fourths is called a **quartile**.

- The median, or second quartile (Q2), splits the data into two halves.
- The first quartile (Q1) is the middle value of the lower half of the data.
- The third quartile (Q3) is the middle value of the upper half of the data.

For example, here is a data set with 11 values.

12	19	20	21	22	33	34	35	40	40	49
		Q1			Q2			Q3		

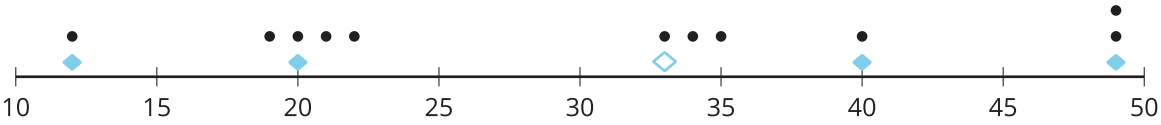
- The median, or second quartile (Q2), splits the data into two equal halves. For this data set, the median is 33.
- The first quartile (Q1) is the middle value of the lower half of the data. For this data set, the first quartile is 20. It is the median of the numbers that are less than 33.
- The third quartile (Q3) is the middle value of the upper half of the data. For this data set, the third quartile is 40. It is the median of the numbers that are greater than 33.

The difference between the maximum and minimum values of a data set is the **range**. For this data set, the range is 37 because $49 - 12 = 37$.

The difference between Q3 and Q1 is the **interquartile range (IQR)**. For this data set, the IQR is 20 because $\text{Q3} - \text{Q1} = 40 - 20$. Because the distance between Q1 and Q3 includes the middle two-fourths of the distribution, the values between those two quartiles are sometimes called the *middle half of the data*.

The bigger the IQR, the more spread out the middle half of the data values are. The smaller the IQR, the closer together the middle half of the data values are. This is why we can use the IQR as a measure of spread.

A *five-number summary* can be used to summarize a distribution. It includes the minimum, first quartile, median, third quartile, and maximum of the data set. For the previous example, the five-number summary is 12, 20, 33, 40, and 49. These numbers are marked with diamonds on the dot plot.



Different data sets can have the same five-number summary. For instance, here is another data set with the same minimum, maximum, and quartiles as the previous example.