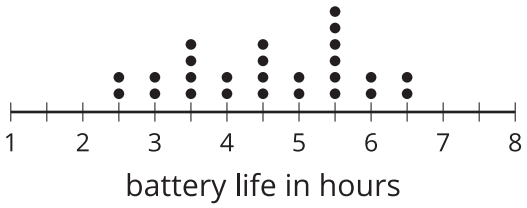
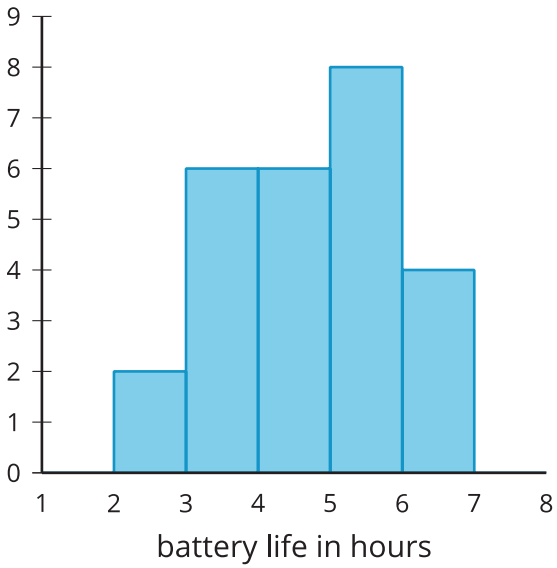
Data Representations

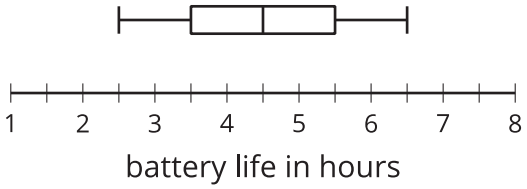
Let's represent and analyze data using dot plots, histograms, and box plots.

2.1 Notice and Wonder: Battery Life

The dot plot, histogram, and box plot summarize the hours of battery life for 26 cell phones that are constantly streaming video. What do you notice? What do you wonder?







2.2 Tomato Plants: Histogram

A histogram can be used to represent the distribution of numerical data.

1. The data represent the number of days it takes for different tomato plants to produce tomatoes. Use the information to complete the frequency table.

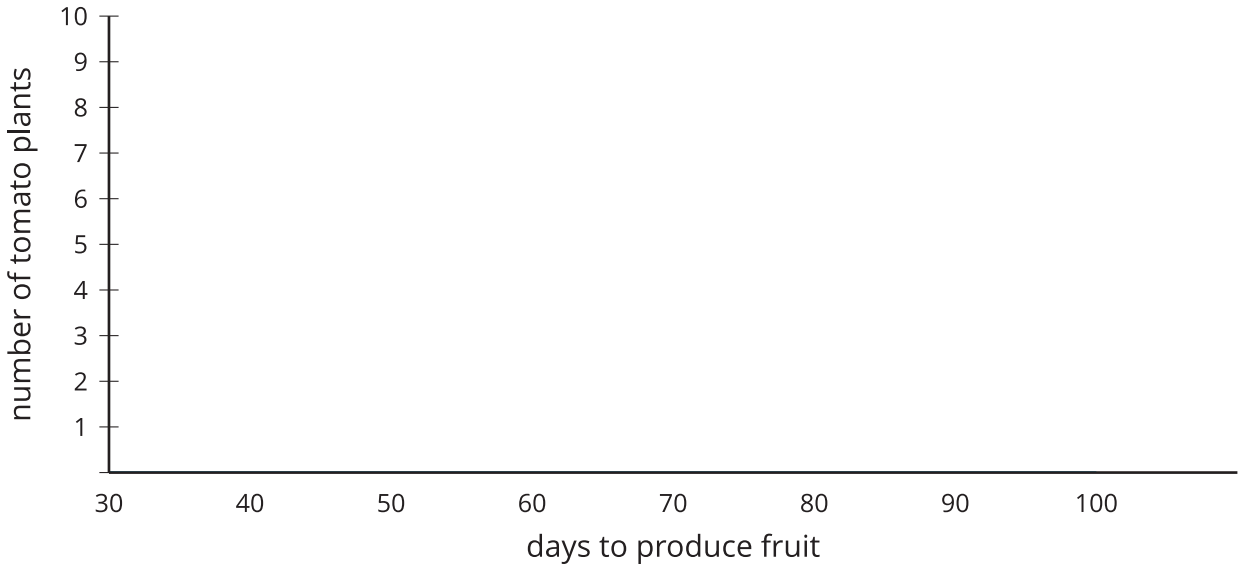
47 52 53 55 57 60 61 62 63 65 65

65 65 68 70 72 72 75 75 75 76 77

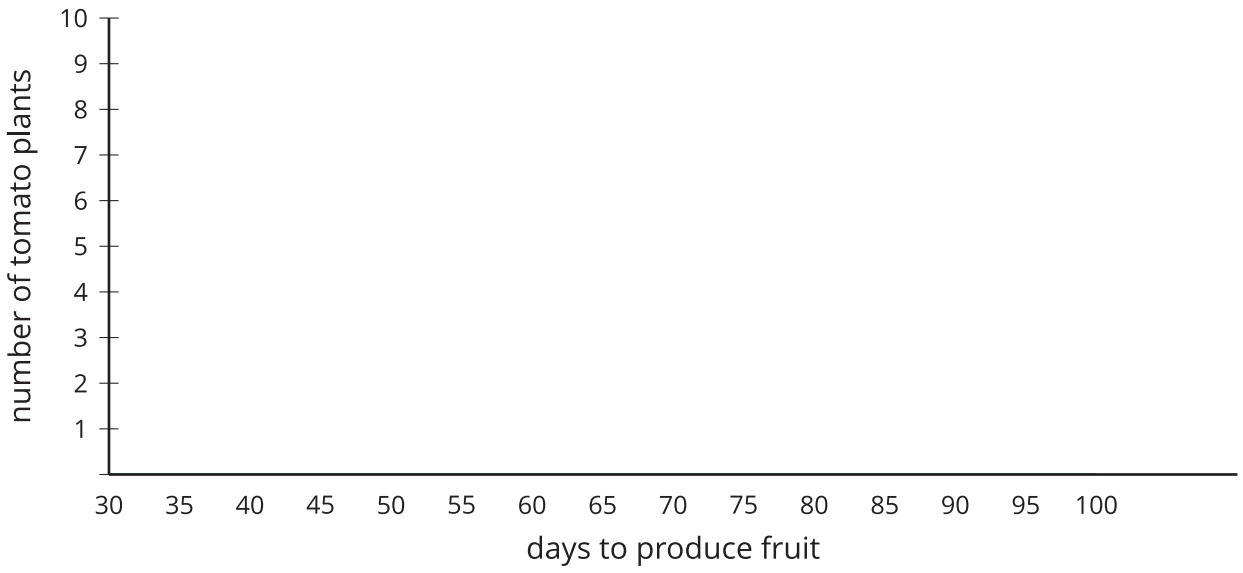
78 80 81 82 85 88 89 90

days to produce fruit	frequency
40–50	
50–60	
60–70	
70–80	
80–90	
90–100	

2. Use the set of axes and the information in your table to create a histogram.



3. The histogram you created has intervals of width 10 (like 40–50 and 50–60). Use the set of axes and data to create another histogram with an interval of width 5. How does this histogram differ from the other one?



Are you ready for more?

It often takes some playing around with the interval widths to figure out which gives the best sense of the shape of the distribution.

1. What might be a problem with using interval widths that are too large?

2. What might be a problem with using interval widths that are too small?

3. What other considerations might go into choosing the width of an interval?

2.3 Tomato Plants: Box Plot

A box plot can also be used to represent the distribution of numerical data.

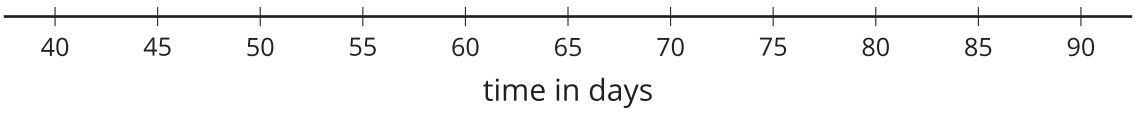
minimum	Q1	median	Q3	maximum

1. Using the same data as in the previous activity for tomato plants, find the median, and add it to the table. What does the median represent for these data?

2. Find the median of the least 15 values to split the data into the first and second quarters. This value is called the first quartile. Add this value to the table under Q1. What does this value mean in this situation?

3. Find the value (the third quartile) that splits the data into the third and fourth quarters, and add it to the table under Q3. Add the minimum and maximum values to the table.

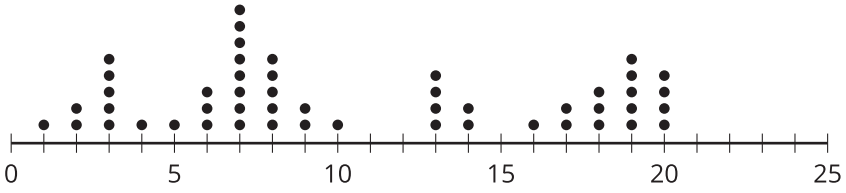
4. Use the **five-number summary** to create a box plot that represents the number of days it takes for these tomato plants to produce tomatoes.



Lesson 2 Summary

The table shows a list of the number of minutes people could intensely focus on a task before needing a break. Fifty people of different ages are represented.

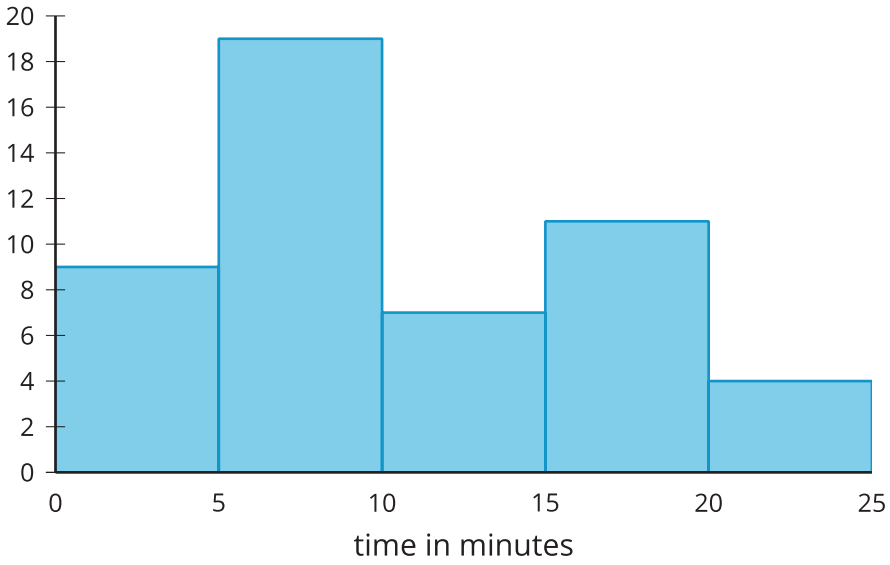
19	7	1	16	20	2	7	19	9	13	3	9	18	13	20	8	3	14	13	2
8	5	17	7	18	17	8	8	7	6	2	20	7	7	10	7	6	19	3	18	8
19	7	13	20	14	6	3	19	4											



In a situation like this, it is helpful to represent the data graphically to better notice any patterns or other interesting features in the data. A dot plot can be used to see the shape and **distribution** of the data.

There were quite a few people that lost focus at around 3, 7, 13, and 19 minutes, and nobody lost focus at 11, 12, or 15 minutes. Dot plots are useful when the data set is not too large and shows all of the individual values in the data set. In this example, a dot plot can easily show all of the data. If the data set is very large (more than 100 values, for example), or if there are many different values that are not exactly the same, it may be hard to see all of the dots on a dot plot.

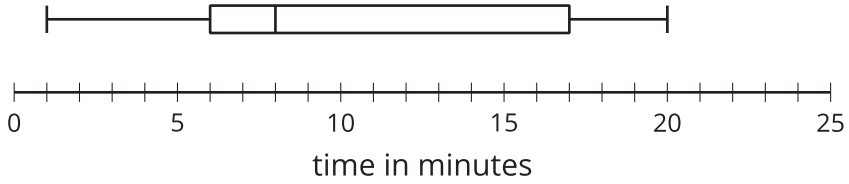
A histogram is another representation that shows the shape and distribution of the same data.



Most people lost focus between 5 and 10 minutes or between 15 and 20 minutes, while only 4 of the 50 people got distracted between 20 and 25 minutes. When creating histograms, each interval includes the number at the lower end of the interval but not the number at the upper end.

For example, the tallest bar displays values that are greater than or equal to 5 minutes but less than 10 minutes. In a histogram, values that are in an interval are grouped together. Although the individual values get lost with the grouping, a histogram can still show the shape of the distribution.

Here is a box plot that represents the same data.



Box plots are created using a **five-number summary**. For a set of data, the five-number summary consists of these five statistics: the minimum value, the first quartile, the median, the third quartile, and the maximum value. These values split the data into four sections, each representing approximately one-fourth of the data. The median of this data is indicated at 8 minutes, and about 25% of the data fall in the short second quarter of the data between 6 and 8 minutes. Similarly, approximately one-fourth of the data are between 8 and 17 minutes. Like the histogram, the box plot does not show individual data values, but other features such as quartiles, range, and median are seen more easily. Dot plots, histograms, and box plots provide three different ways to look at the shape and distribution while highlighting different aspects of the data.