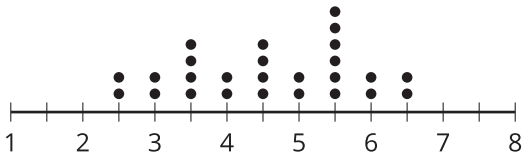
Representaciones de datos

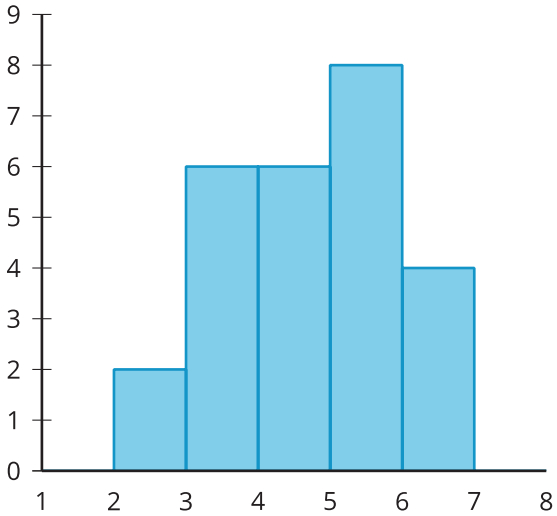
Representemos y analicemos datos usando diagramas de puntos, histogramas y diagramas de caja.

2.1 Observa y pregúntate: Duración de la batería

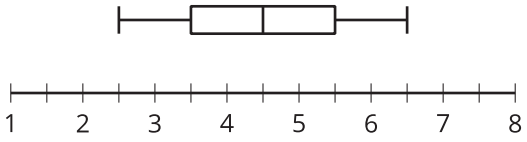
El diagrama de puntos, el histograma y el diagrama de caja resumen cuántas horas duran las baterías de 26 teléfonos celulares que reproducen videos de internet constantemente. ¿Qué observas? ¿Qué te preguntas?



duración de la batería en horas



duración de la batería en horas



duración de la batería en horas

2.2 Plantas de tomate: Histograma

Se puede usar un histograma para representar la distribución de datos numéricos.

1. Los datos representan el número de días que ciertas plantas de tomate tardan en producir tomates. Usa la información para completar la tabla de frecuencias.

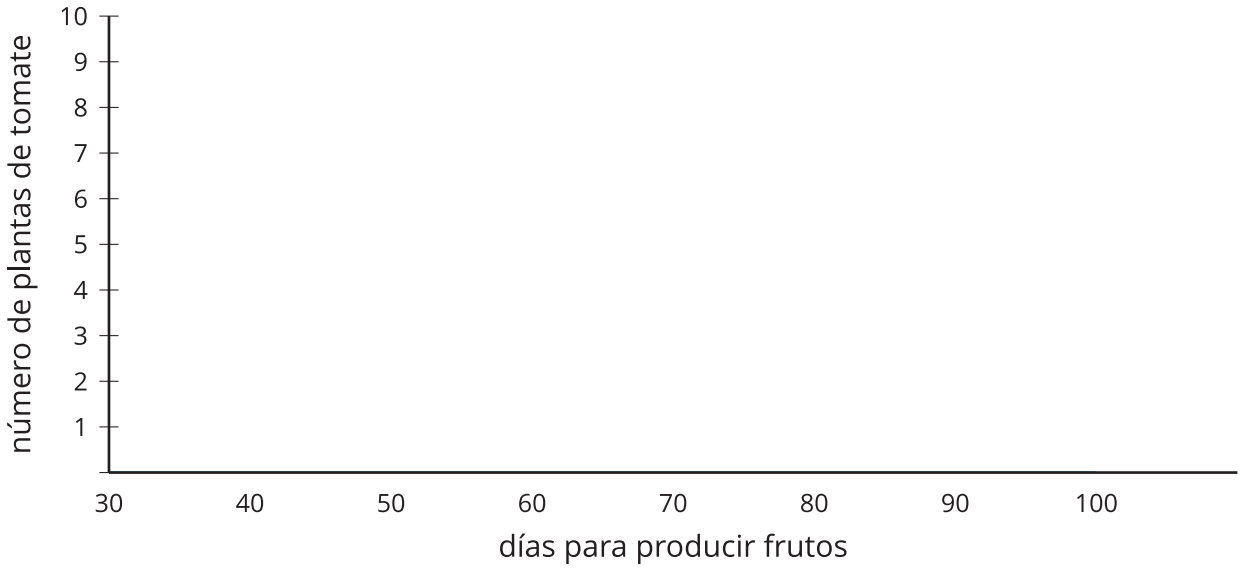
47 52 53 55 57 60 61 62 63 65 65

65 65 68 70 72 72 75 75 75 76 77

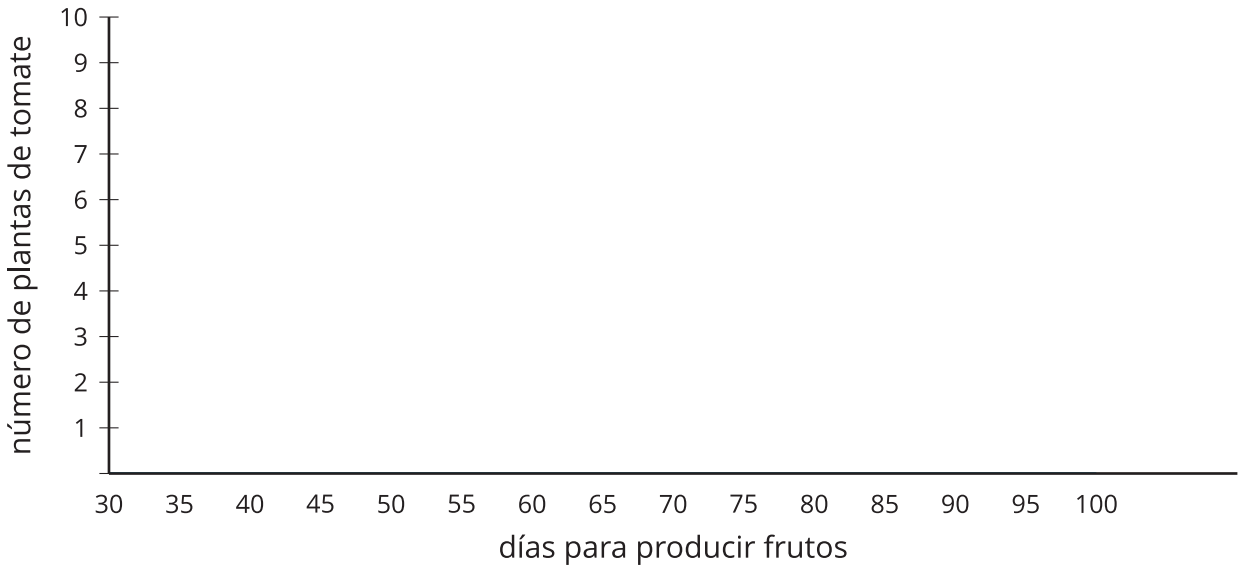
78 80 81 82 85 88 89 90

días para producir frutos	frecuencia
40–50	
50–60	
60–70	
70–80	
80–90	
90–100	

2. Usa los ejes y la información de tu tabla para crear un histograma.



3. El histograma que creaste tiene intervalos de 10 de ancho (por ejemplo, de 40 a 50 y de 50 a 60). Usa los ejes y los datos para crear otro histograma que tenga intervalos de 5 de ancho. ¿En qué son diferentes este histograma y el otro?



¿Estás listo para más?

A menudo conviene probar intervalos de distintos anchos para determinar cuál nos da una mejor idea de la forma de la distribución.

1. ¿Cuál podría ser un problema si usáramos intervalos de anchos muy grandes?

2. ¿Cuál podría ser un problema si usáramos intervalos de anchos muy pequeños?

3. ¿Qué otras consideraciones se deben tener en cuenta al escoger el ancho de un intervalo?

2.3 Plantas de tomate: Diagrama de caja

También se puede usar un diagrama de caja para representar la distribución de datos numéricos.

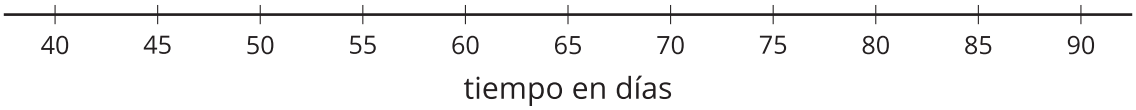
mínimo	Q1	mediana	Q3	máximo

1. Usando los mismos datos de las plantas de tomate que usaste en la actividad anterior, encuentra la mediana y ponla en la tabla. ¿Qué representa la mediana en el caso de estos datos?

2. Encuentra la mediana de los 15 valores más pequeños para separar los datos en el primer y el segundo cuarto. Este valor se llama el primer cuartil. Pon este valor en la tabla debajo de la casilla Q1. ¿Qué significa este valor en esta situación?

3. Encuentra el valor que separa los datos en el tercer y el cuarto cuarto (el tercer cuartil) y ponlo en la tabla debajo de la casilla Q3. Escribe el valor mínimo y el valor máximo en la tabla.

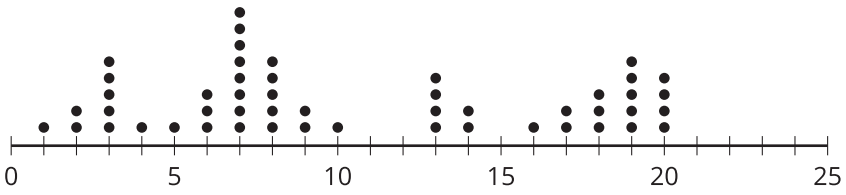
4. Usa el **resumen de cinco números** para crear un diagrama de caja que represente el número de días que tardan estas plantas de tomate en producir tomates.



Resumen de la lección 2

La tabla muestra una lista del número de minutos que las personas se pueden concentrar intensamente en una tarea antes de tener que descansar. Cincuenta personas de diversas edades están representadas en la tabla.

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											

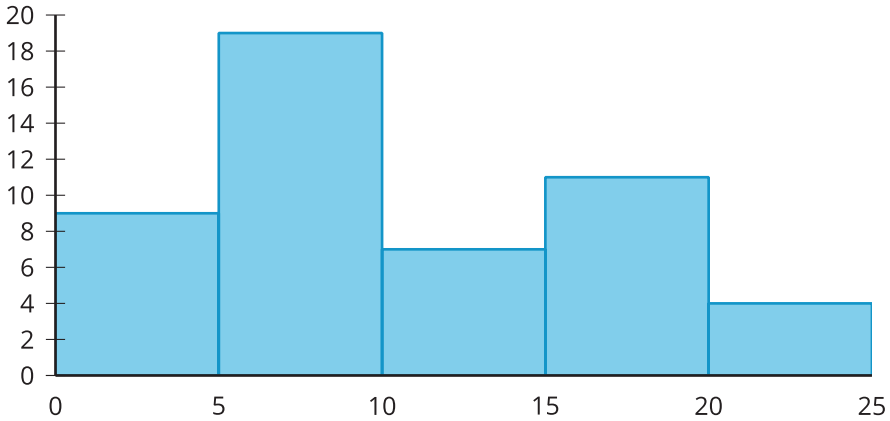


tiempo en minutos

En una situación como esta, es útil representar los datos de manera gráfica para observar mejor los patrones u otras características interesantes de los datos. Se puede usar un diagrama de puntos para ver la forma y la **distribución** de los datos.

Hubo varias personas que perdieron la concentración cerca de los 3, los 7, los 13 y los 19 minutos, y nadie perdió la concentración a los 11 ni a los 12 ni a los 15 minutos. Los diagramas de puntos son útiles cuando el conjunto de datos no es muy grande y se muestran todos los valores individuales del conjunto de datos. En este ejemplo, un diagrama de puntos puede mostrar fácilmente todos los datos. Si el conjunto de datos es muy grande (por ejemplo, si tiene más de 100 valores) o si hay muchos valores que no son exactamente el mismo, puede ser difícil ver todos los puntos en un diagrama de puntos.

Un histograma es otra representación que muestra la forma y la distribución de los mismos datos.

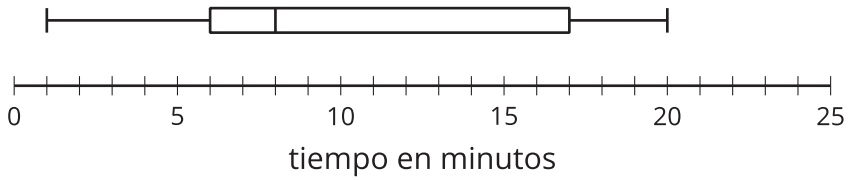


tiempo en minutos

La mayoría de las personas perdió la concentración cuando pasaron entre 5 y 10 minutos o entre 15 y 20 minutos, mientras que solo 4 de las 50 personas se distrajeron cuando habían pasado entre 20 y 25 minutos. Al crear histogramas debemos tener en cuenta que cada intervalo incluye el número del extremo inferior del intervalo, pero no el del extremo superior.

Por ejemplo, la barra más alta muestra valores que son mayores o iguales a 5 minutos y menores que 10 minutos. En un histograma, los valores que están en un intervalo están agrupados. Aunque los valores individuales se pierden al ser agrupados, un histograma también puede mostrar la forma de la distribución.

Este es un diagrama de caja que representa los mismos datos.



Los diagramas de caja se crean usando un **resumen de cinco números**. El resumen de cinco números de un conjunto de datos consiste en estos cinco estadísticos: el valor mínimo, el primer cuartil, la mediana, el tercer cuartil y el valor máximo. Estos valores separan los datos en cuatro secciones y cada una representa aproximadamente un cuarto de los datos. La mediana de estos datos está marcada en 8 minutos, y aproximadamente el 25% de los datos está en el segundo cuarto de los datos, entre 6 y 8 minutos. De manera similar, aproximadamente un cuarto de los datos está entre 8 y 17 minutos. Al igual que el histograma, el diagrama de caja no muestra los valores individuales de los datos, pero se pueden ver más fácilmente otras características, como los cuartiles, el rango y la mediana. Los diagramas de puntos, los histogramas y los diagramas de caja nos muestran la forma y la distribución de los datos de tres maneras distintas y resaltan aspectos distintos de los datos.