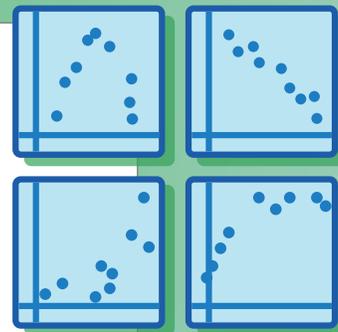


Unidad **6**

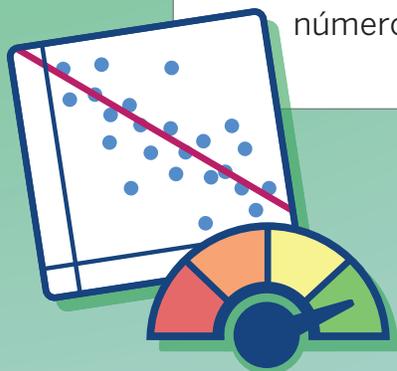
# Asociación de datos



Los datos sirven para comprender el mundo que nos rodea. Organizar y presentar datos nos permite describir tendencias y hacer predicciones. En esta unidad, aprenderás que hay puntos de datos que representan dos elementos de información.

## Preguntas esenciales

- ¿Qué es un diagrama de dispersión y qué información puede revelar?
- ¿Cómo pueden ayudar las rectas a modelar datos en un diagrama de dispersión?
- ¿Cómo puedes analizar datos con dos variables que son categorías en lugar de números?



Puedes organizar y presentar datos que incluyen números de diferentes maneras, tales como una tabla y un diagrama de dispersión. Un **diagrama de dispersión** es un conjunto de puntos de datos desconectados, trazados en un plano de coordenadas.

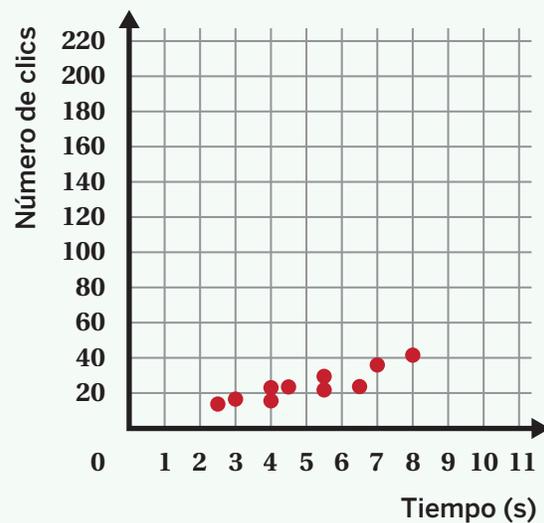
Una tabla y un diagrama de dispersión muestran los mismos datos, pero pueden ser útiles de diferentes maneras. Por ejemplo, puedes utilizar un diagrama de dispersión para analizar las conexiones entre dos variables, mientras que una tabla es útil para buscar los valores exactos de puntos de datos específicos.

Aquí se presenta un conjunto de datos que muestra la cantidad de tiempo, en segundos, y el número de clics realizados.

Tabla

Tiempo (s)	Número de clics
2.5	14
3	17
4	16
4	23
4.5	24
5.5	22
5.5	30
6.5	24
7	36
8	42

Diagrama de dispersión



## Prueba a hacer esto

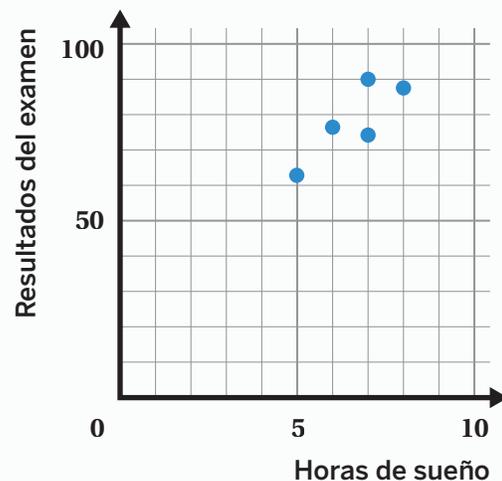
Para obtener créditos adicionales, un grupo de estudiantes participó en un estudio en el que registraron el número de horas que durmieron la noche anterior a un examen y sus resultados en el mismo.

Este es un diagrama de dispersión de los datos.

- Ayaan: 7 horas, puntuación de 74
- Emika: 6 horas, puntuación de 76
- Inola: 8 horas, puntuación de 88
- Kwasi: 5 horas, puntuación de 63
- Zoe: 7 horas, puntuación de 90

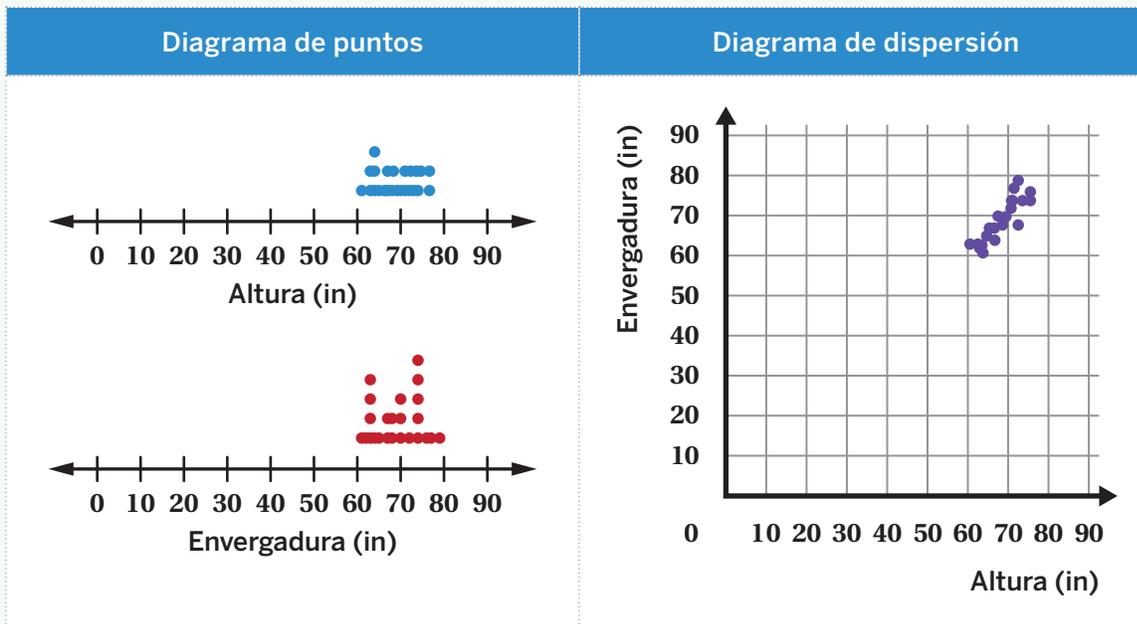
**a** ¿De qué otra forma podrías organizar o mostrar estos datos?

**b** ¿Cuál es una ventaja de organizar los datos en un diagrama de dispersión?



Los datos que se representan como números, cantidades o medidas y que pueden compararse de forma significativa se denominan *datos numéricos* o *datos cuantitativos*. Puedes analizar *datos univariados*, que implican una variable, y *datos bivariados*, que implican dos variables.

Existen diferentes formas de representar datos numéricos. Un *diagrama de puntos* muestra los datos de una variable y un *diagrama de dispersión* muestra los datos de dos variables al mismo tiempo. Ver dos variables numéricas al mismo tiempo nos permite observar tendencias y conexiones.

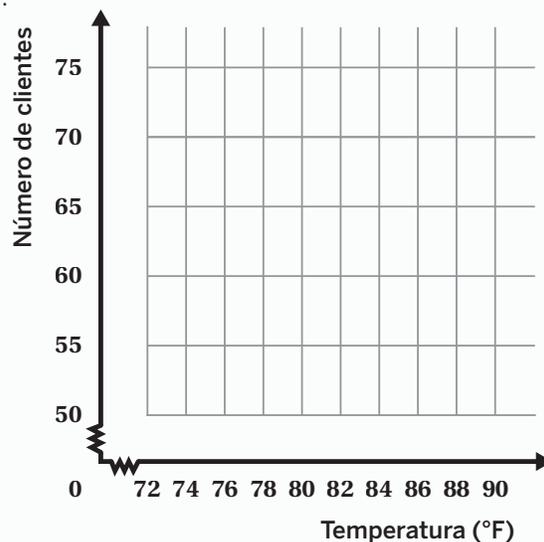


## Prueba a hacer esto

Durante una semana, un puesto de helados recopiló datos sobre la temperatura exterior y el número de clientes.

- a** Crea un diagrama de dispersión con estos datos.

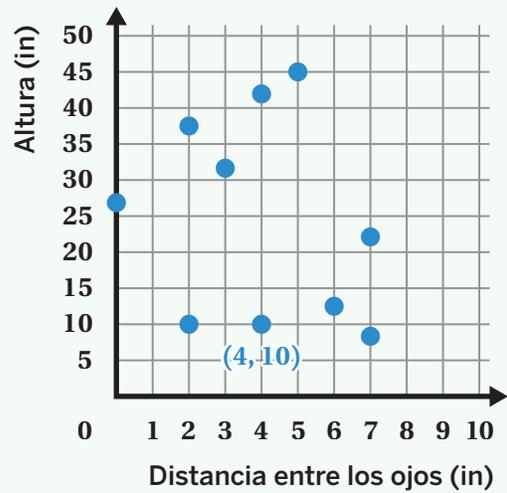
Día	Temperatura (°F)	Número de clientes
Lunes	85	58
Martes	83	55
Miércoles	90	68
Jueves	75	50
Viernes	85	72



- b** Escribe una pregunta que puedas responder basándote en el diagrama de dispersión.

Un punto en un diagrama de dispersión representa dos elementos de información. Los nombres de los ejes indican cómo interpretar las coordenadas de cada punto.

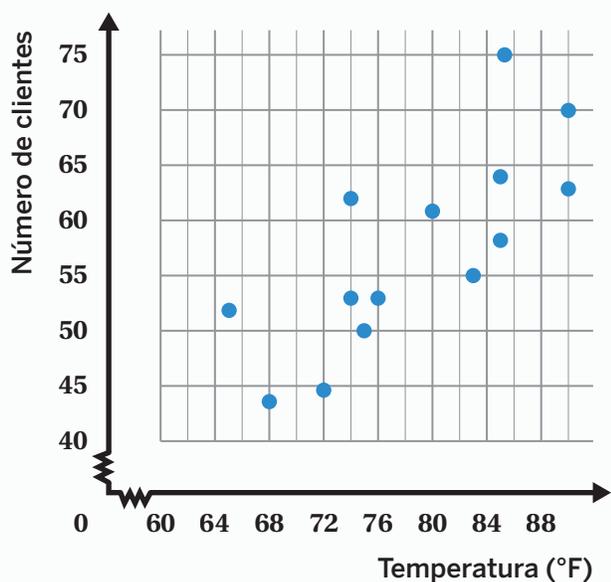
En este ejemplo, el punto (4, 10) representa un robot con una distancia entre los ojos de 4 pulgadas y una altura de 10 pulgadas.



## Prueba a hacer esto

Un puesto de helados recopiló datos sobre la temperatura exterior y el número de clientes a lo largo del tiempo.

- a** Encierra en un círculo el punto que representa el día en el que la temperatura alcanzó los 72°F en el exterior.
- b** Estima el número de clientes de ese día.
- c** ¿Qué representa el punto (85, 75)?

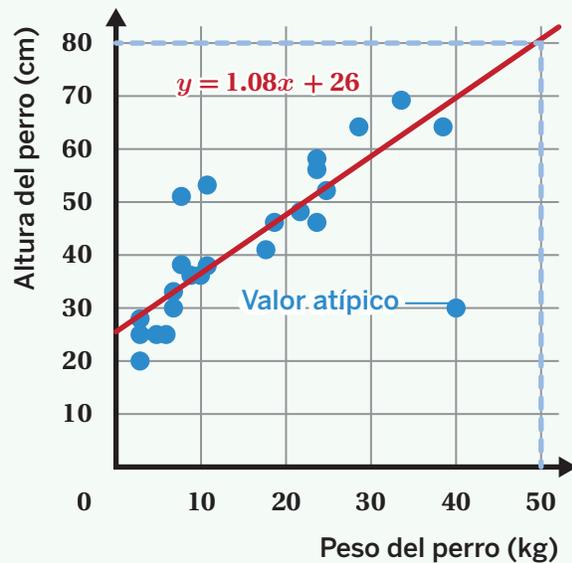


Un **modelo lineal** es una recta en un diagrama de dispersión que ayuda a identificar tendencias en los datos con mayor claridad. También puedes utilizar un modelo lineal para hacer una predicción.

Por ejemplo, hay dos formas de utilizar este modelo lineal para predecir la altura de un perro cuando pesa 50 kilogramos.

- Usa la gráfica para ubicar el 50 en el eje  $x$  y sube hasta encontrarte con el modelo lineal, que indica un valor de  $y$  de 80. Esto significa que cuando el peso del perro es de 50 kilogramos, su altura es de 80 centímetros.
- Usa la ecuación del modelo lineal,  $y = 1.08x + 26$ , sustituyendo la  $x$  por 50 y evaluando  $y$ , lo que da aproximadamente 80 centímetros.

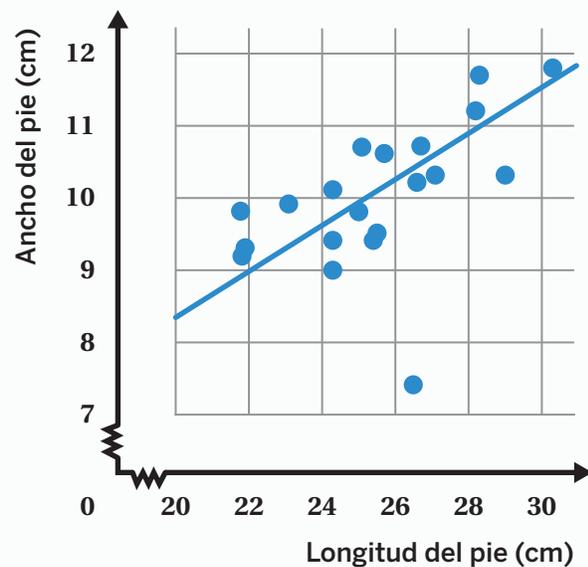
Puedes identificar un **valor atípico** buscando puntos que estén alejados de los demás y de los valores previstos. El punto (40, 30) es un valor atípico en la gráfica de pesos y alturas de los perros.



## Prueba a hacer esto

Este diagrama de dispersión muestra los datos recopilados sobre la longitud y el ancho de los pies de unas personas.

- Usa el modelo lineal para predecir el ancho de los pies de una persona cuya longitud de pies es de 28 centímetros.
- Encierra en un círculo el valor atípico en la gráfica.



Puedes usar un diagrama de dispersión para identificar patrones en un conjunto de datos y relaciones entre dos variables.

Por ejemplo, este diagrama de dispersión muestra datos sobre el tiempo de espera de los clientes en un restaurante *drive-thru* y el número de empleados que trabajaban en ese momento.

El diagrama de dispersión muestra tanto información específica como tendencias generales, incluido lo siguiente:

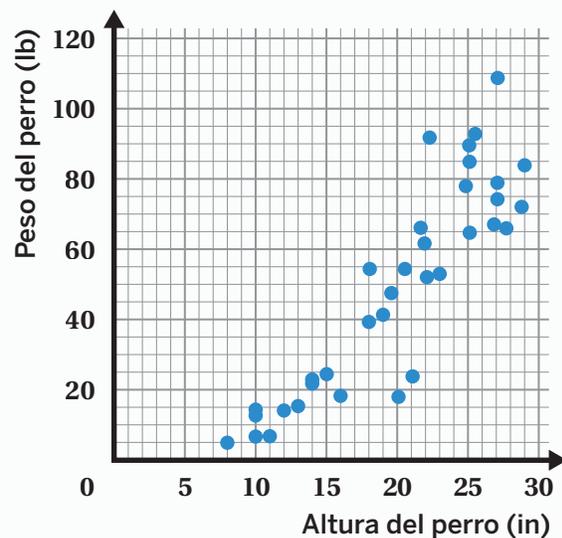
- Cuando había 3 empleados, el tiempo de espera era de aproximadamente 12 minutos.
- Cuanto más personal hay, más corto parece ser el tiempo de espera.



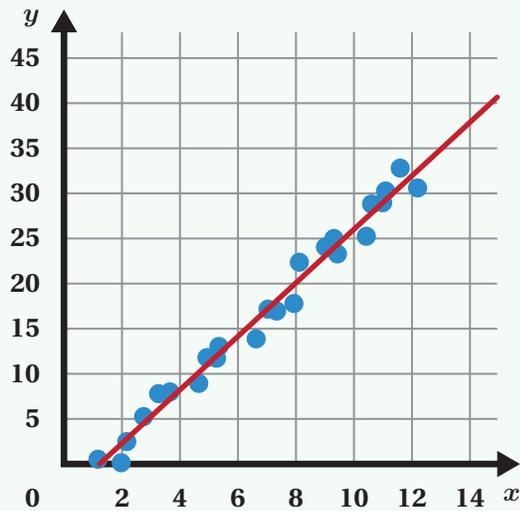
## Prueba a hacer esto

Este diagrama de dispersión muestra las alturas y pesos de 35 perros.

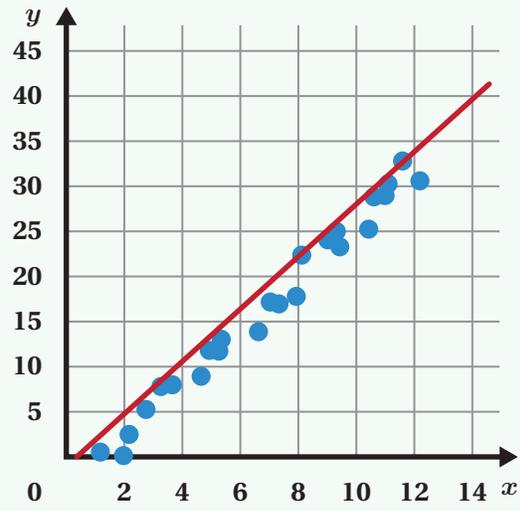
- ¿Qué representa el punto (8, 5) en esta situación?
- Identifica una tendencia general basándote en el diagrama de dispersión.



Al crear una línea de ajuste para un diagrama de dispersión, es importante determinar qué tan bien se ajusta la línea a los datos. Una buena línea de ajuste sigue la tendencia de los datos, está lo más cerca posible de los puntos trazados y tiene aproximadamente el mismo número de puntos por encima y por debajo de la línea. La línea puede pasar por algunos, todos o ninguno de los puntos.



Esta línea se ajusta bien a los datos.

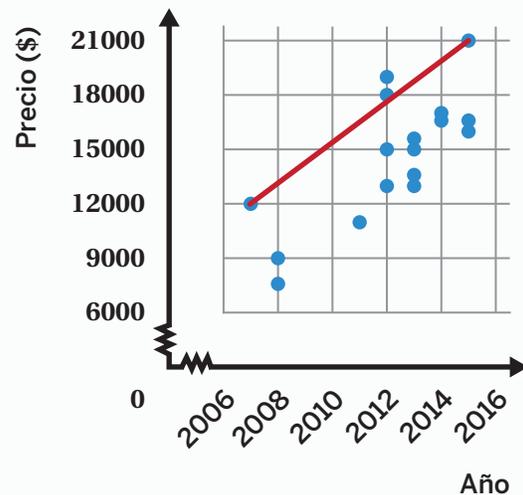


Esta línea no se ajusta bien a los datos.

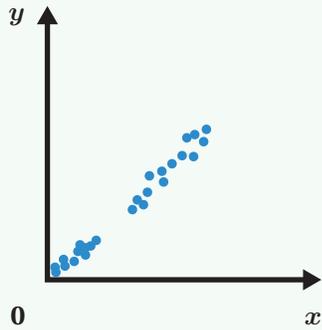
## Prueba a hacer esto

Estos son los datos sobre el precio de un automóvil usado y el año en que se fabricó.

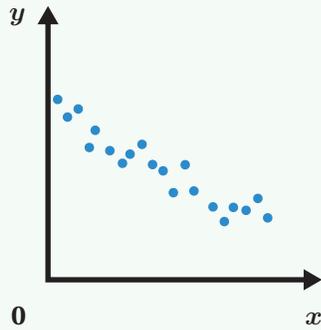
- a** Explica por qué este modelo no se ajusta bien a los datos.
  
- b** Dibuja una línea de ajuste que represente mejor los datos de este diagrama de dispersión.



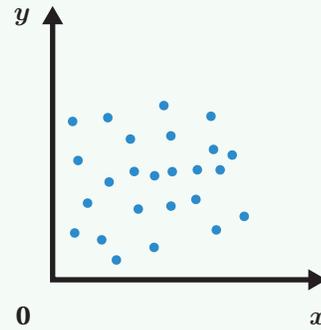
Si dos variables en un diagrama de dispersión están relacionadas, existe una **asociación** entre ellas. La pendiente de un modelo lineal puede ayudar a determinar el tipo de asociación. Una asociación positiva significa que cuando el valor de una variable crece, el de la otra también crece. Una asociación negativa significa que cuando el valor de una variable crece, el de la otra decrece. Si el diagrama de dispersión no muestra una tendencia clara entre las dos variables, entonces las variables no tienen asociación.



Asociación positiva



Asociación negativa

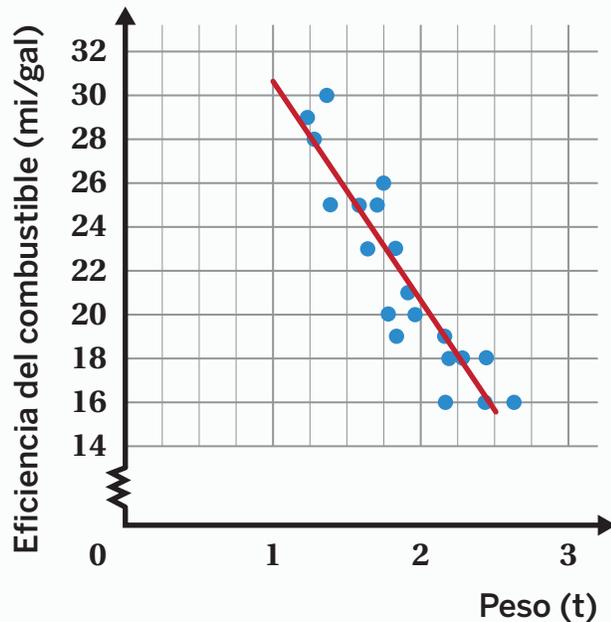


Sin asociación

## Prueba a hacer esto

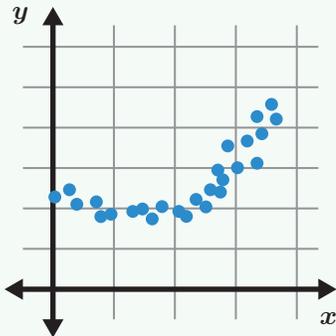
Este es un diagrama de dispersión que muestra datos sobre el peso y el rendimiento del combustible (millas recorridas por galón de combustible) de 21 automóviles.

- a ¿Qué tipo de asociación existe entre el peso y el rendimiento del combustible?
  
- b La línea de ajuste tiene una pendiente de aproximadamente  $-10$ . ¿Qué significa este número en el contexto del peso de un automóvil y su rendimiento de combustible previsto?

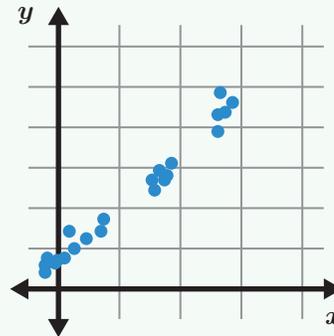


Cuando los datos en un diagrama de dispersión pueden modelarse con una línea recta, decimos que tienen una asociación lineal. Los datos que no se pueden ajustar a una línea recta tienen una asociación no lineal. A veces, ciertos puntos de datos aparecen muy juntos, lo que se denomina **agrupamiento**.

Este diagrama de dispersión es un ejemplo de una asociación no lineal sin agrupamientos.



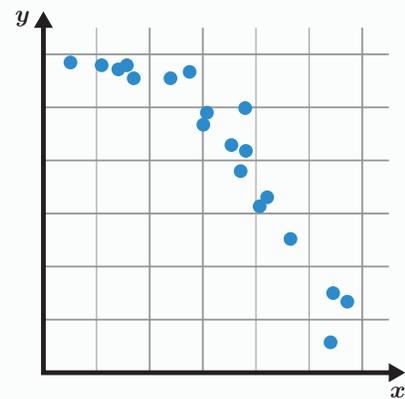
Este diagrama de dispersión es un ejemplo de una asociación lineal con agrupamientos.



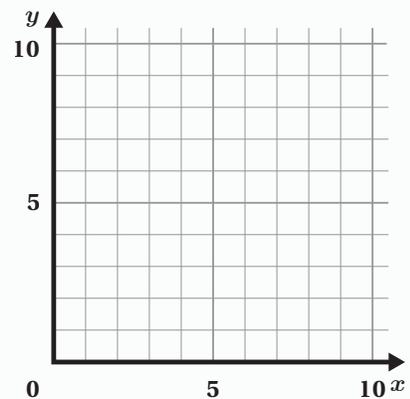
## Prueba a hacer esto

- a** Describe este diagrama de dispersión utilizando al menos dos términos del banco de palabras.

asociación positiva	asociación negativa	agrupamientos
asociación lineal	asociación no lineal	valor atípico



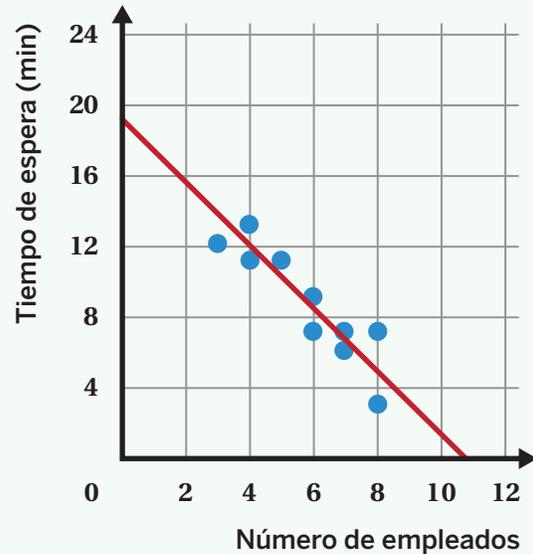
- b** Traza puntos para crear un diagrama de dispersión que muestre una asociación lineal negativa con agrupamientos.



Comprendiendo la asociación entre dos variables se pueden hacer predicciones sobre valores desconocidos. Cuando hay una asociación lineal, usar un modelo lineal suele ayudar a que las predicciones sean más precisas.

Por ejemplo, este diagrama de dispersión muestra datos sobre el tiempo de espera, en minutos, de los clientes en un restaurante *drive-thru* y el número de empleados que estaban trabajando en ese momento. Estos datos se pueden modelar mediante la ecuación  $y = -1.75x + 19$ .

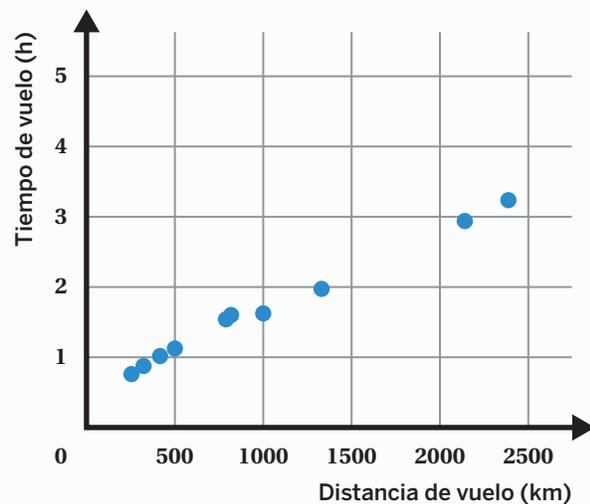
- La *pendiente* del modelo lineal es  $-1.75$ , lo que significa que si el número de empleados aumenta en 1, el tiempo de espera disminuye en 1.75 minutos.
- El modelo lineal predice que, cuando hay 2 empleados trabajando, el tiempo de espera será de aproximadamente 15.5 minutos.
- Pero el modelo lineal también predice que, cuando hay 0 empleados trabajando, el tiempo de espera sería de 19 minutos, ¡lo que es imposible!



## Prueba a hacer esto

Este diagrama de dispersión muestra las distancias y los tiempos de una serie de vuelos.

- Dibuja la línea de ajuste que mejor representa los datos.
- Describe la asociación entre la distancia de vuelo y el tiempo de vuelo.
- Utiliza tu modelo para predecir el tiempo de vuelo cuando la distancia de vuelo es de 2,000 kilómetros.



Puedes utilizar una **tabla de doble entrada** para comparar dos variables de *datos categóricos*, que son datos que pueden clasificarse en categorías. Las tablas de doble entrada muestran una de las variables en la parte superior y la otra en uno de los lados. Cada entrada de la tabla representa la **frecuencia**, o el número de veces que una categoría aparece en el conjunto de datos.

Por ejemplo, esta tabla de doble entrada muestra datos que indican si los estudiantes meditaron un día determinado y si ese día se sintieron relajados o agitados.

	Meditaron	No meditaron	Total
Relajados	45	8	53
Agitados	23	21	44
Total	68	29	97

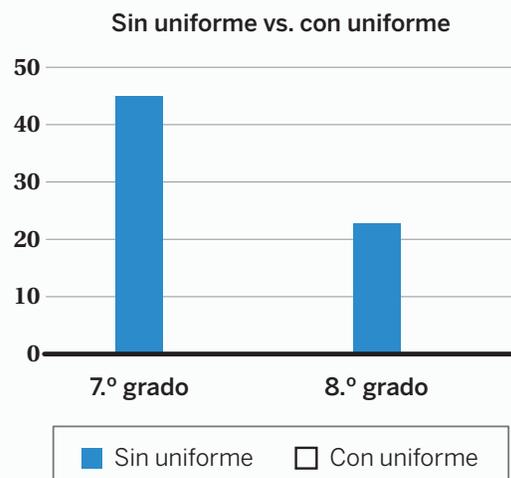
Se pueden utilizar tablas de doble entrada, diagramas de dispersión y diagramas de barras para representar datos y explorar asociaciones entre estos datos. Cada tipo de representación tiene sus ventajas y desventajas. Puedes utilizar estas representaciones para investigar posibles conexiones entre variables. En el ejemplo, observamos que existe una conexión entre meditar y sentirse relajados, ya que la mayoría de las personas que se sentían relajadas también habían meditado.

## Prueba a hacer esto

Los estudiantes hicieron una encuesta sobre si querían o no querían llevar uniforme. La tabla muestra sus respuestas.

	Sin uniforme	Con uniforme	Total
7.º grado	45		53
8.º grado	23	21	
Total		29	97

- a** Completa la tabla de doble entrada.
- b** ¿Qué significa el número 53 en esta situación?
- c** Completa el diagrama de barras de esta situación con los datos de los estudiantes que quieren uniformes.



Puedes usar tipos específicos de tablas de doble entrada y diagramas de barras para mostrar frecuencias y porcentajes dentro de conjuntos de datos.

La **frecuencia relativa** de una categoría es la fracción o el porcentaje del conjunto de datos que pertenece a esa categoría. Una tabla de doble entrada con frecuencias relativas muestra la fracción o el porcentaje de cada categoría en lugar del número de puntos de datos.

**Frecuencias relativas**

	Prefiere mensajes	Prefiere llamadas	Total
Menor de 40 años	82%	18%	100%
40 años o más	33%	67%	100%

Un **diagrama de barras segmentadas** compara diferentes categorías dentro de un conjunto de datos. Cada barra representa todos los datos dentro de una categoría, o el 100%. Las barras están divididas en partes o segmentos, que representan el porcentaje de cada subcategoría dentro de la categoría.

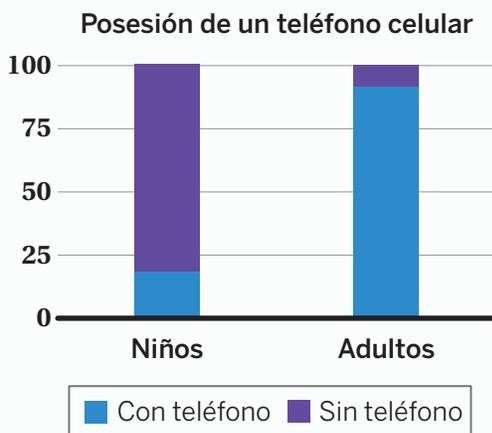
Diagrama de barra segmentadas



Podemos usar representaciones como estas para identificar asociaciones entre dos variables categóricas. Por ejemplo, la tabla y el diagrama muestran una asociación entre las variables categóricas de edad y preferencia de comunicación.

## Prueba a hacer esto

Aquí se muestran dos representaciones de frecuencia relativa.



**Calcetines de la suerte y ganadores de bingo**

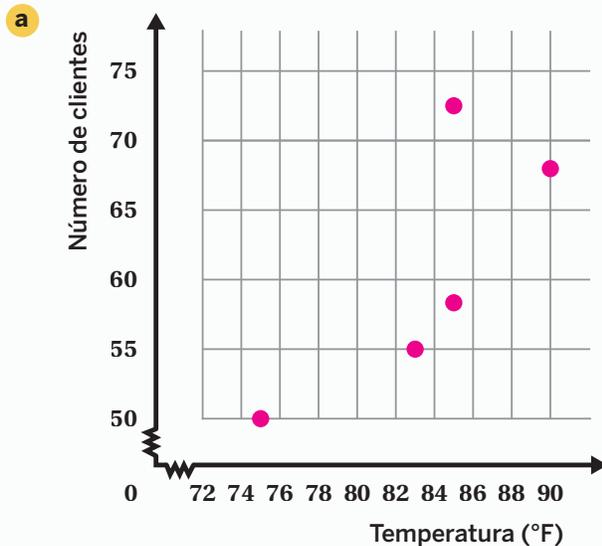
	Ganadores	Perdedores	Total
Calcetines de la suerte	80%	20%	100%
Calcetines normales	79%	21%	100%

- a** ¿Existe alguna asociación entre la edad y la posesión de un teléfono celular? Explica tu razonamiento.
- b** ¿Existe alguna asociación entre usar calcetines de la suerte y ganar al bingo? Explica tu razonamiento.

## Lección 1

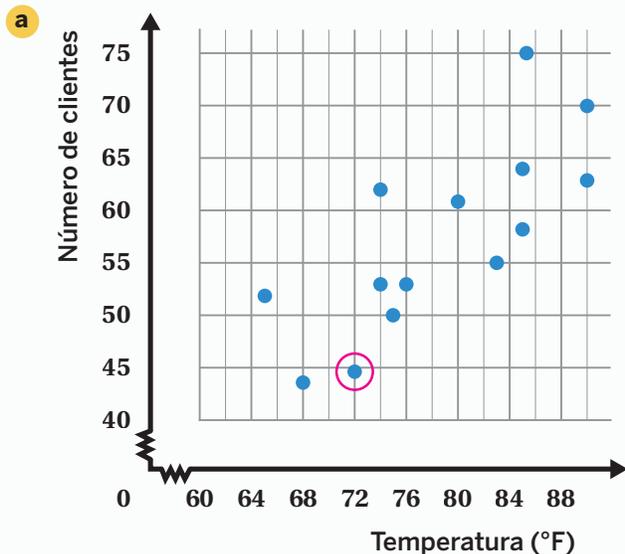
- a *Las respuestas pueden variar. Crea una tabla que ordene los datos por horas de sueño o puntuación.*
- b *Las respuestas pueden variar. En un diagrama de dispersión, es más fácil ver la relación entre las dos variables.*

## Lección 2



- b *Las respuestas pueden variar. ¿Cómo se relacionan la temperatura y el número de clientes? ¿Qué tendencias se observan en el diagrama de dispersión?*

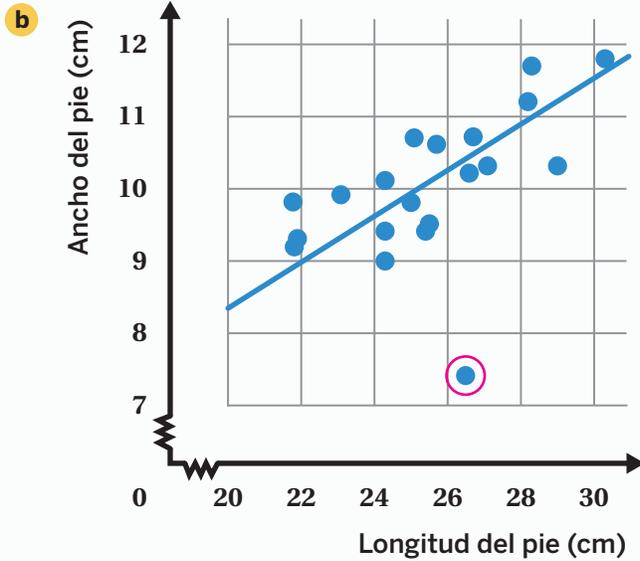
## Lección 3



- b *Había aproximadamente 44 clientes cuando la temperatura exterior era de 72°F.*
- c *El punto (85, 75) representa una temperatura de 85°F en un día con 75 clientes.*

### Lección 4

a Las respuestas entre 10.75 y 10.95 centímetros se consideran correctas.



### Lección 5

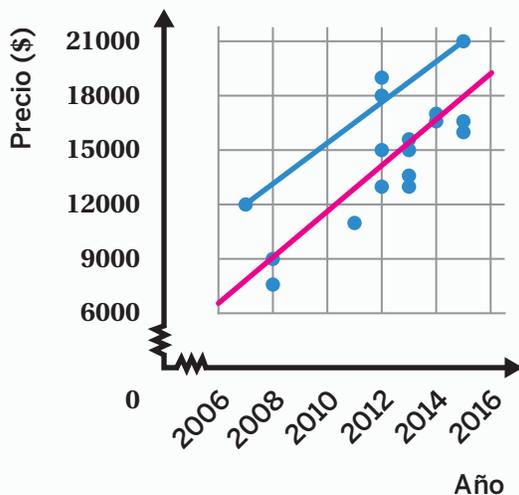
a El punto (8, 5) representa un perro que mide 8 pulgadas y pesa 5 libras.

b Según el diagrama de dispersión, la tendencia general es que los perros más altos pesan más.

### Lección 6

a Las respuestas pueden variar. Este modelo no se ajusta bien porque hay más datos por debajo que por encima de la línea. Esto significa que la línea sobreestimaría el valor de la mayoría de los datos.

b La gráfica muestra un ejemplo.

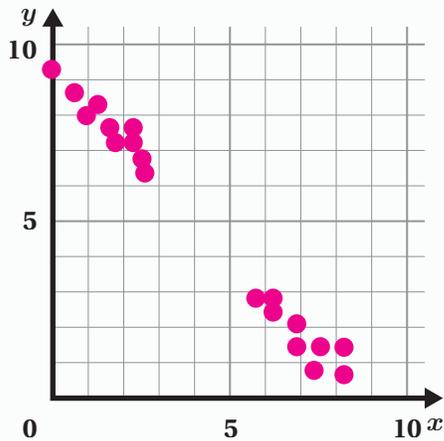


## Lección 7

- a *Las respuestas pueden variar.* Existe una asociación negativa entre el peso y el rendimiento del combustible. A medida que aumenta el peso del automóvil, disminuye la eficiencia de combustible.
- b Si el peso de un automóvil aumenta en 1 tonelada, el modelo predice que el rendimiento del combustible del automóvil disminuirá en 10 millas por galón.

## Lección 8

- a *Las respuestas pueden variar.* Este diagrama de dispersión muestra una asociación no lineal negativa sin agrupamientos.
- b *Las respuestas pueden variar.* La gráfica muestra un ejemplo.



## Lección 9

- a 

A scatter plot with a grid. The horizontal axis is labeled 'Distancia de vuelo (km)' and has tick marks at 0, 500, 1000, 1500, 2000, and 2500. The vertical axis is labeled 'Tiempo de vuelo (h)' and has tick marks at 1, 2, 3, 4, and 5. There are 10 data points showing a clear positive linear trend. A pink line of best fit is drawn through the points, starting at approximately (0, 0.5) and ending at approximately (2500, 3.5).
- b *Las respuestas pueden variar.* La distancia de vuelo y el tiempo de vuelo tienen una asociación lineal positiva, lo que significa que a medida que aumenta la distancia de vuelo, aumenta el tiempo de vuelo.
- c *Las respuestas pueden variar.* Un vuelo con una distancia de 2,000 kilómetros tendrá una duración de poco menos de 3 horas.

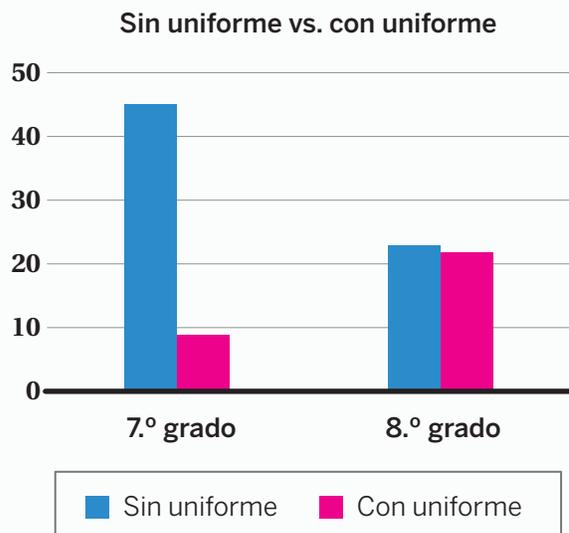
### Lección 10

a

	Sin uniforme	Con uniforme	Total
7.º grado	45	8	53
8.º grado	23	21	44
Total	68	29	97

b En esta situación, 53 representa el número total de estudiantes de 7.º grado que respondieron la encuesta.

c



### Lección 11

- a **Sí.** Las explicaciones pueden variar. Existe una asociación entre la edad y la posesión de teléfonos celulares. El porcentaje de adultos que poseen un teléfono celular es mucho mayor que el porcentaje de niños que poseen un teléfono celular.
- b **No.** Las explicaciones pueden variar. No hay ninguna relación entre usar calcetines de la suerte y ganar al bingo. El porcentaje de ganadores de bingo que usan calcetines de la suerte es solo un 1% mayor que el porcentaje de ganadores de bingo que usan calcetines normales.

## English

## Español

### A

**association** If two variables are related to one another, we say that there is an association between them. Associations can be described as positive or negative, linear or non-linear.

Some examples of associations are:

Positive: One variable increases as the other also increases.

Negative: One variable decreases as the other increases.

Linear: Can be modeled by a straight line.

Non-linear: Cannot be modeled by a straight line.

**asociación** Si dos variables están relacionadas entre sí, decimos que existe una asociación entre ambas. Las asociaciones pueden describirse como positivas o negativas, lineales o no lineales.

Algunos ejemplos de asociaciones son los siguientes:

Positiva: El valor de una variable aumenta a medida que el valor de la otra también aumenta. Negativa: El valor de una variable disminuye a medida que el valor de la otra aumenta.

Lineal: Puede modelarse con una línea recta.

No lineal: No puede modelarse con una línea recta.

### B

**bivariate data** Data that involves two variables. Each data point contains two pieces of information.

For example, a collection of students' heights and shoe sizes would be a bivariate data set.

**datos bivariados** Datos en los que se incluyen dos variables. Cada punto de datos contiene dos informaciones.

Por ejemplo, una colección de estaturas de estudiantes y tamaños de zapatos sería un conjunto de datos bivariados.

### C

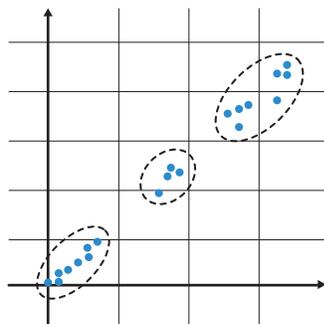
**categorical data** Data that can be sorted into categories instead of counted.

For example, data on a group of students' favorite colors or what month they were born in.

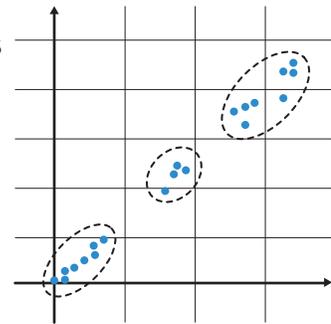
**datos categóricos** Datos que pueden clasificarse en categorías en lugar de contarse.

Por ejemplo, los datos de los colores favoritos de un grupo de estudiantes o el mes en el que nacieron.

**clusters** Groups of data values that are close together.

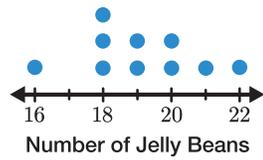


**agrupaciones (clusters)** Grupos de valores de datos que están próximos entre sí.



## English

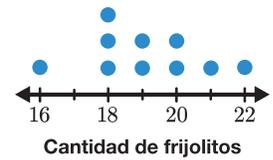
**dot plot** A way to visualize numerical data sets, where each data point is represented by a dot on a number line. Data points with the same value are stacked on top of each other. A dot plot is sometimes called a line plot.



For example, this dot plot shows that 3 students guessed that there were 18 jelly beans in a jar.

## Español

**diagrama de puntos** Una forma de visualizar conjuntos de datos numéricos, en la que cada punto de datos se representa mediante un punto en una recta numérica. Los puntos de datos con el mismo valor se apilan unos sobre otros. Un diagrama de puntos a veces se conoce como gráfica de puntos.



Por ejemplo, este diagrama de puntos muestra que 3 estudiantes estimaron que había 18 frijolitos de jalea en un tarro.

## D

## F

**frequency** The number of times a value appears in a data set.

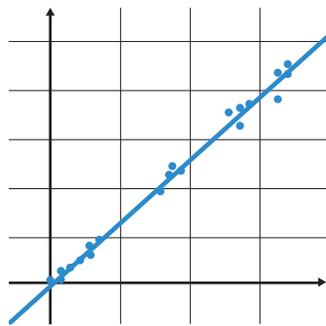
**frecuencia** El número de veces que aparece un valor en un conjunto de datos.

## L

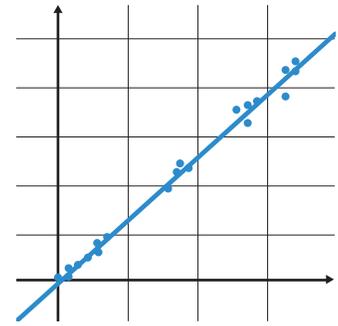
**line of fit** See *linear model*.

**línea de ajuste** Ver *modelo lineal*.

**linear model** A line that shows, or models, the general direction or trend of a group of points in a data set. We can use linear models to make predictions about values related to a given data set.



**modelo lineal** Una línea que muestra, o modela, la dirección o tendencia general de un grupo de puntos en un conjunto de datos. Podemos emplear modelos lineales para hacer predicciones sobre valores relativos a un conjunto de datos determinado.



## N

**numerical data** Data presented as numbers, quantities, or measurements that can be meaningfully compared. It is sometimes called quantitative data.

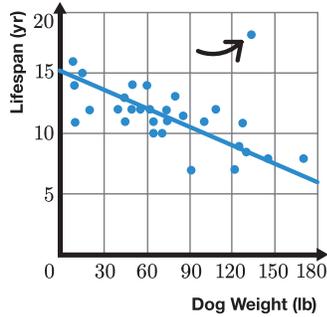
For example, data on a group of students' heights or arm spans.

**datos numéricos** Datos que se presentan como números, cantidades o medidas que pueden compararse de forma significativa. A veces se denominan datos cuantitativos.

Por ejemplo, los datos de las estaturas o las longitudes de los brazos de un grupo de estudiantes.

## English

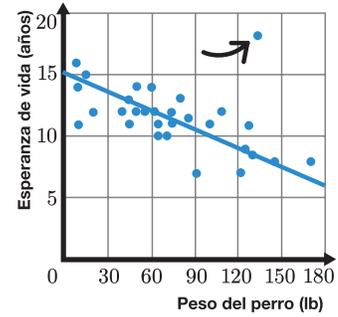
**outlier** A data value that is far from the other values in the data set.



## O

**valor atípico** Un valor que está lejos de los demás valores del conjunto de datos.

## Español



## Q

**quantitative data** See *numerical data*.

**datos cuantitativos** Ver *datos numéricos*.

## R

### relative frequency

The fraction or percent of the data that is in a category instead of the actual number of data points.

	Meditated	Did Not Meditate
Calm	66%	28%
Agitated	34%	72%
Total	100%	100%

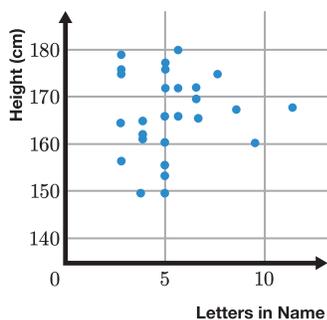
### frecuencia relativa

La fracción o el porcentaje de los datos que pertenecen a una categoría en lugar de la cantidad real de puntos de datos.

	Meditaron	No meditaron
Calmados	66%	28%
Agitados	34%	72%
Total	100%	100%

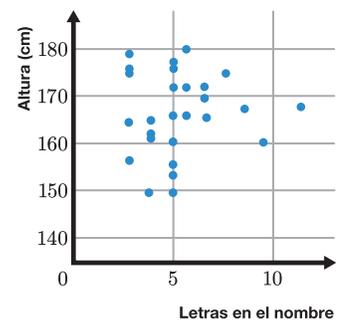
## S

**scatter plot** A set of disconnected data points plotted on a coordinate plane. Scatter plots allow us to investigate connections between two variables.



### diagrama de dispersión

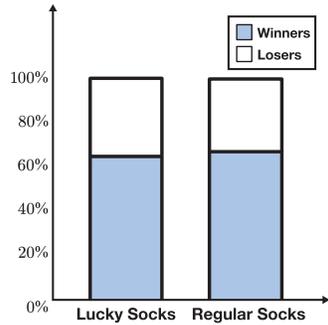
Un conjunto de puntos de datos que no están conectados trazados en un plano de coordenadas. Los diagramas de dispersión nos permiten analizar las conexiones entre dos variables.



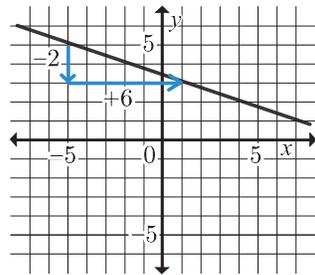
English

**segmented bar graph**

A graph that compares two categories within a data set. The whole bar represents all the data within one category. Then each bar is separated into parts (segments) that show the percentage of each part in the second category.



**slope** A number that describes the direction and steepness of a line. Slope represents the amount that  $y$  changes when  $x$  increases by 1. That's why the slope of a line is sometimes called a rate of change. To calculate the slope, divide the vertical distance between any two points on the line by the horizontal distance between those points.

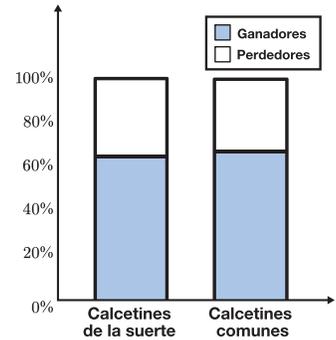


The slope of this line is  $-\frac{2}{6} = -\frac{1}{3}$ .

Español

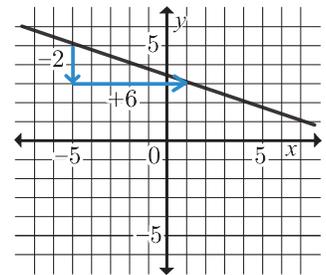
**diagrama de barras segmentadas**

Una gráfica que compara dos categorías dentro de un conjunto de datos. Una barra completa representa todos los datos dentro de una categoría. Luego, cada barra se separa en partes (segmentos) para mostrar el porcentaje de cada parte en la segunda categoría.



**pendiente**

Un número que describe la dirección e inclinación de una línea. La pendiente representa la cantidad en la que cambia  $y$  cuando  $x$  se incrementa en 1. Es por eso que la pendiente de una recta a veces se denomina tasa de cambio. Para calcular la pendiente, la distancia vertical entre dos puntos cualesquiera en la recta se divide entre la distancia horizontal entre dichos puntos.



La pendiente de esta recta es  $-\frac{2}{6} = -\frac{1}{3}$ .

T

**two-way table**

A way to compare two categorical variables. The table shows one of the variables across the top and the other variable down one side. Each entry is the frequency or relative frequency of the category in that column and row.

	Meditated	Did Not Meditate	Total
Calm	45	8	53
Agitated	23	21	44
Total	68	29	97

**tabla de doble entrada**

Una manera de comparar dos variables categóricas. La tabla muestra una de las variables en la parte superior y la otra variable a un costado. Cada entrada es la frecuencia o frecuencia relativa de la categoría que se muestra en esa columna y fila.

	Meditaron	No meditaron	Total
Calmadros	45	8	53
Agitados	23	21	44
Total	68	29	97

# Grade 8 Unit 6 Glossary/8.º grado Unidad 6 Glosario

## English

## Español

### U

**univariate data** A data set that involves one variable. Each data point contains one piece of information.

A collection of students' heights is a univariate data set.

Tables, dot plots, and bar graphs are useful for displaying univariate data.

**datos univariados** Un conjunto de datos que incluye una variable. Cada punto de datos contiene una información.

Una colección de las estaturas de estudiantes es un conjunto de datos univariado.

Las tablas, los diagramas de puntos y los diagramas de barras son útiles para mostrar datos univariados.

### Y

**y-intercept** See *vertical intercept*.

**intersección con el eje  $y$**  Véase *intersección vertical*.