Lesson 6: Scatter Plots

Example 1

A bivariate data set consists of observations on two variables. For example, you might collect data on 13 different car models. Each observation in the data set would consist of an (x, y) pair.

x = weight (in pounds, rounded to the nearest 50 pounds)

and

y = fuel efficiency (in miles per gallon, mpg.)

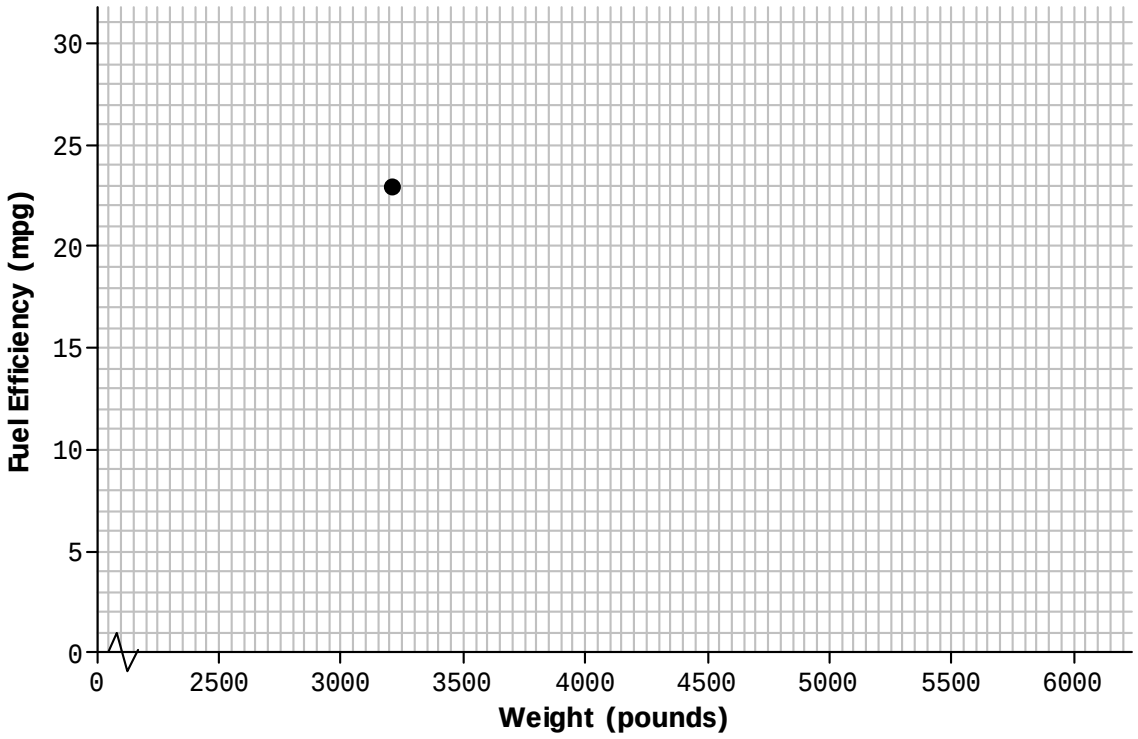
The table below shows the weight and fuel efficiency for 13 car models with automatic transmissions manufactured in 2009 by Chevrolet.

Model	Weight (pounds)	Fuel Efficiency (mpg)
1	3,200	23
2	2,550	28
3	4,050	19
4	4,050	20
5	3,750	20
6	3,550	22
7	3,550	19
8	3,500	25
9	4,600	16
10	5,250	12
11	5,600	16
12	4,500	16
13	4,800	15

Exercises 1–3

1. In the table above, the observation corresponding to model 1 is $(3200, 23)$. What is the fuel efficiency of this car? What is the weight of this car?

2. Add the points corresponding to the other 12 observations to the scatter plot.



3. Do you notice a pattern in the scatter plot? What does this imply about the relationship between weight (x) and fuel efficiency (y)?

Exercises 4–8

Is there a relationship between price and the quality of athletic shoes? The data in the table below are from the Consumer Reports website.

x = price (in dollars)

and

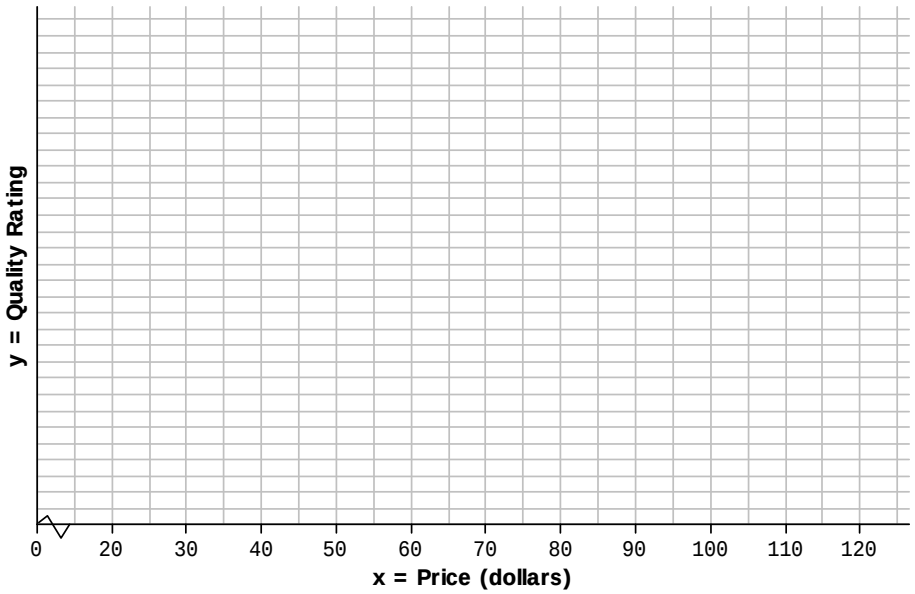
y = Consumer Reports quality rating

The quality rating is on a scale of 0 to 100, with 100 being the highest quality.

Shoe	Price (dollars)	Quality Rating
1	65	71
2	45	70
3	45	62
4	80	59
5	110	58
6	110	57
7	30	56
8	80	52
9	110	51
10	70	51

4. One observation in the data set is $(110, 57)$. What does this ordered pair represent in terms of cost and quality?

5. To construct a scatter plot of these data, you need to start by thinking about appropriate scales for the axes of the scatter plot. The prices in the data set range from \$30 to \$110, so one reasonable choice for the scale of the x-axis would range from \$20 to \$120, as shown below. What would be a reasonable choice for a scale for the y-axis?



6. Add a scale to the y-axis. Then, use these axes to construct a scatter plot of the data.

7. Do you see any pattern in the scatter plot indicating that there is a relationship between price and quality rating for athletic shoes?

8. Some people think that if shoes have a high price, they must be of high quality. How would you respond?

Example 2: Statistical Relationships

A pattern in a scatter plot indicates that the values of one variable tend to vary in a predictable way as the values of the other variable change. This is called a *statistical relationship*. In the fuel efficiency and car weight example, fuel efficiency tended to decrease as car weight increased.

This is useful information, but be careful not to jump to the conclusion that increasing the weight of a car *causes* the fuel efficiency to go down. There may be some other explanation for this. For example, heavier cars may also have bigger engines, and bigger engines may be less efficient. You cannot conclude that changes to one variable *cause* changes in the other variable just because there is a statistical relationship in a scatter plot.

Exercises 9–10

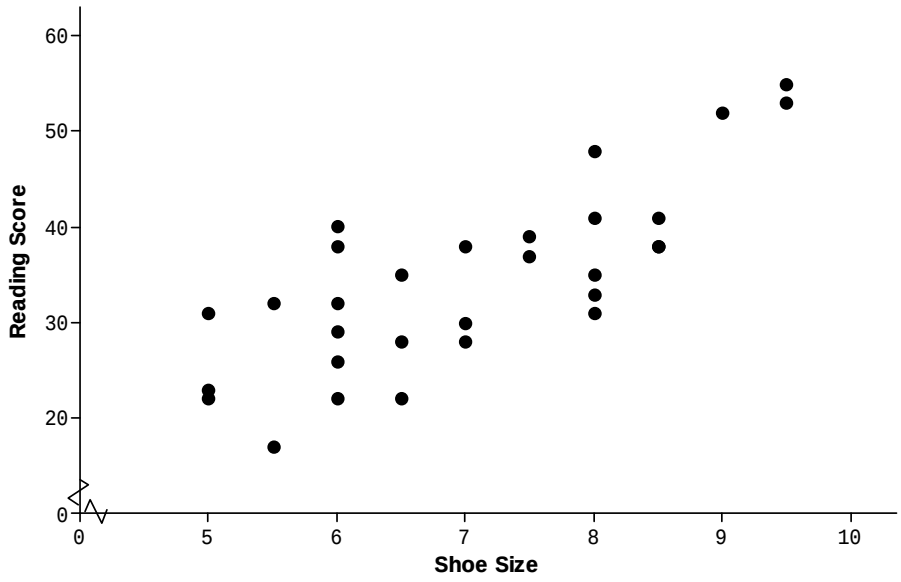
9. Data were collected on

   x = shoe size

   and

   y = score on a reading ability test

   for 30 elementary school students. The scatter plot of these data is shown below. Does there appear to be a statistical relationship between shoe size and score on the reading test?



10. Explain why it is not reasonable to conclude that having big feet causes a high reading score. Can you think of a different explanation for why you might see a pattern like this?

Lesson Summary

- A scatter plot is a graph of numerical data on two variables.
- A pattern in a scatter plot suggests that there may be a relationship between the two variables used to construct the scatter plot.
- If two variables tend to vary together in a predictable way, we can say that there is a statistical relationship between the two variables.
- A statistical relationship between two variables does not imply that a change in one variable causes a change in the other variable (a cause-and-effect relationship).

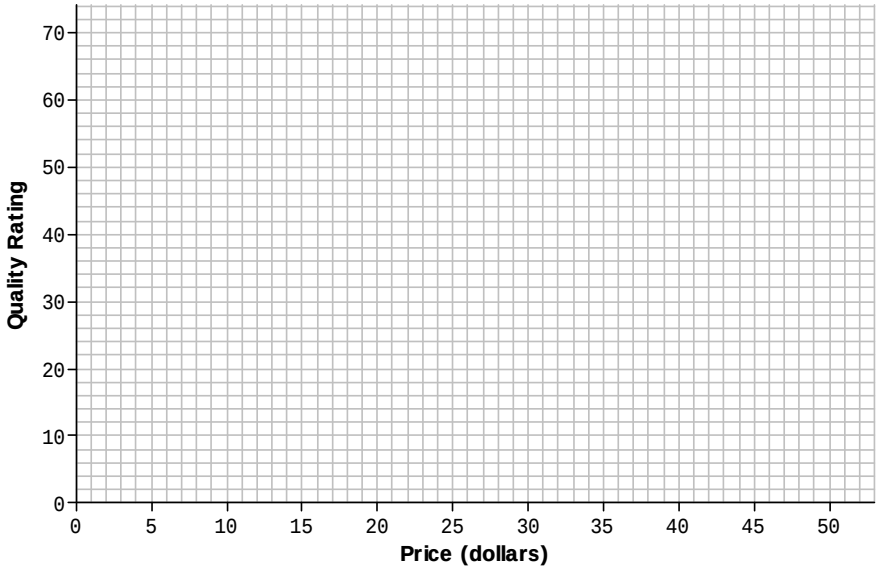
Problem Set

1. The table below shows the price and overall quality rating for 15 different brands of bike helmets.

   Data Source: www.consumerreports.org

Helmet	Price (dollars)	Quality Rating
A	35	65
B	20	61
C	30	60
D	40	55
E	50	54
F	23	47
G	30	47
H	18	43
I	40	42
J	28	41
K	20	40
L	25	32
M	30	63
N	30	63
O	40	53

   Construct a scatter plot of price (x) and quality rating (y). Use the grid below.



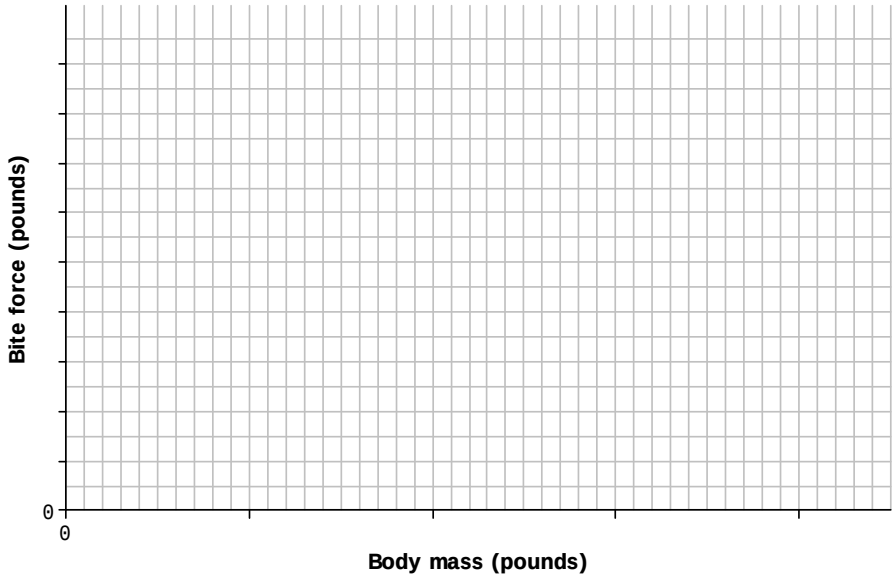
2. Do you think that there is a statistical relationship between price and quality rating? If so, describe the nature of the relationship.

3. Scientists are interested in finding out how different species adapt to finding food sources. One group studied crocodiles to find out how their bite force was related to body mass and diet. The table below displays the information they collected on body mass (in pounds) and bite force (in pounds).

Species	Body mass (pounds)	Bite force (pounds)
Dwarf crocodile	35	450
Crocodile F	40	260
Alligator A	30	250
Caiman A	28	230
Caiman B	37	240
Caiman C	45	255
Croc A	110	550
Nile crocodile	275	650
Croc B	130	500
Croc C	135	600
Croc D	135	750
Caiman D	125	550
Indian Gharial croc	225	400
Crocodile G	220	1,000
American Croc	270	900
Croc D	285	750
Croc E	425	1,650
American Alligator	300	1,150
Alligator B	325	1,200
Alligator C	365	1,450

Data Source: PLoS One Greg Erickson biomechanics, Florida State University

Construct a scatter plot of body mass (x) and bite force (y). Use the grid below, and be sure to add an appropriate scale to the axes.



4. Do you think that there is a statistical relationship between body mass and bite force? If so, describe the nature of the relationship.

5. Based on the scatter plot, can you conclude that increased body mass causes increased bite force? Explain.