Lesson 7: Observing more patterns in scatter plots

Goals

- Categorise data sets, and describe (orally) the properties used to create categories.
- Describe (orally) features of data on scatter plots, including “linear” and “nonlinear association” and “clustering” using informal language.
- Explain (orally) what might cause a nonlinear association or clustering of data points in context. Association is used here as correlation has not yet been formally defined.

Learning Targets

- I can pick out clusters in data from a scatter plot.
- I can use a scatter plot to decide if two variables have a linear association.

Lesson Narrative

In this lesson, students see non-linear associations for the first time and do a card sort to distinguish linear and non-linear associations. They visually identify clusters in data. Students will not study non-linear associations or clustering using quantitative tools; instead they will rely on visual patterns in scatter plots.

Addressing

- Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

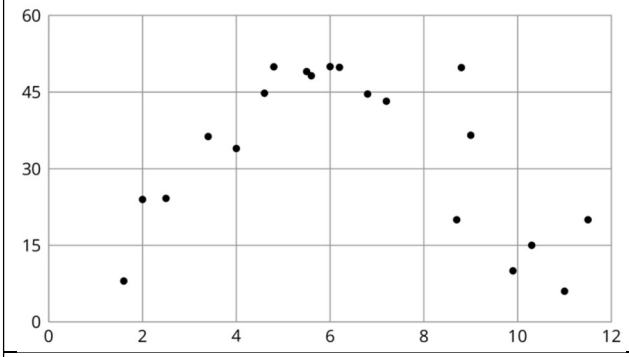
Instructional Routines

- Collect and Display
- Discussion Supports
- Notice and Wonder
- Take Turns
- Think Pair Share

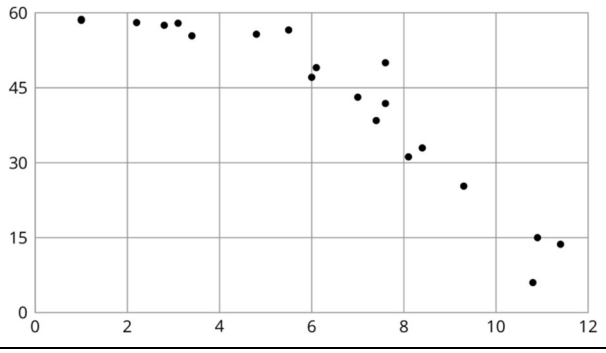
Required Materials

Pre-printed cards, cut from copies of the blackline master

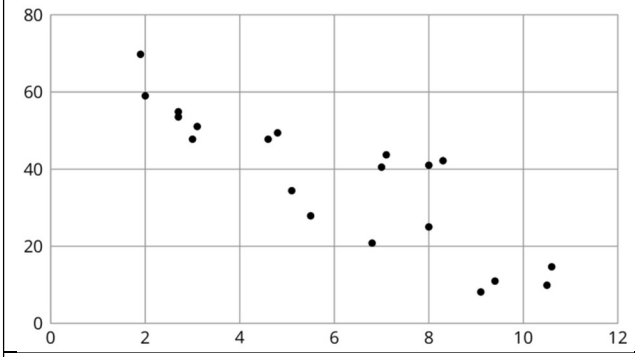
Scatter Plot City
A



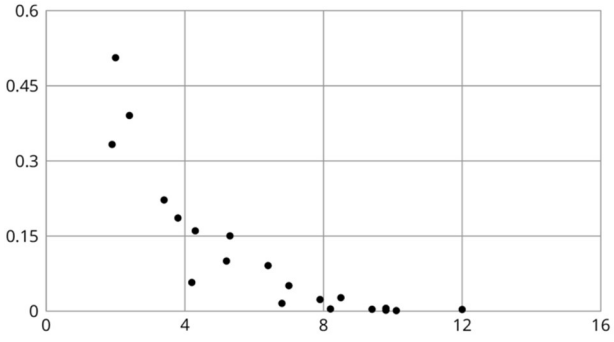
Scatter Plot City
B



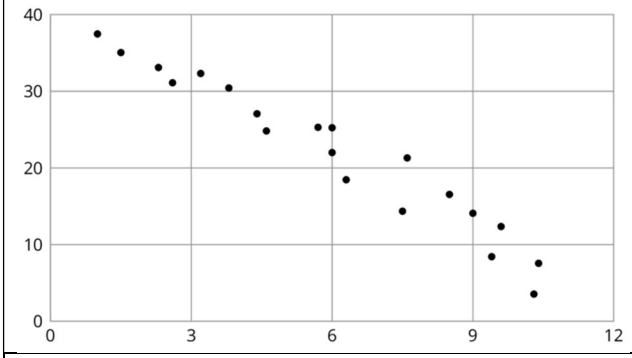
Scatter Plot City
C



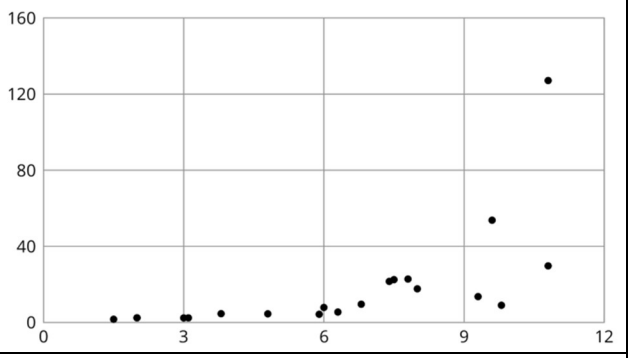
Scatter Plot City
D



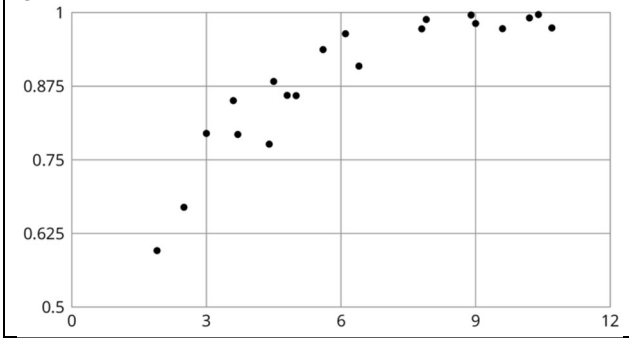
Scatter Plot City
E



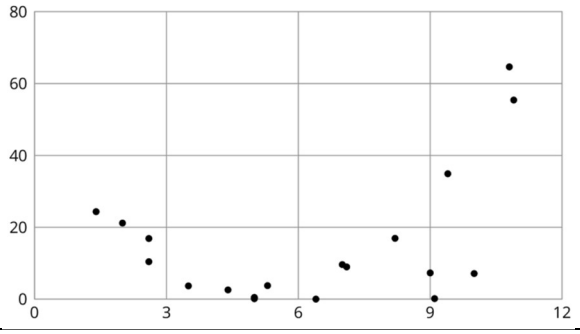
Scatter Plot City
F



Scatter Plot City
G



Scatter Plot City
H



Required Preparation

Print and cut up cards from the Scatter Plot City blackline master. Prepare 1 set of cards for every student.

Student Learning Goals

Let's look for other patterns in data.

7.1 Notice and Wonder: Nonlinear Scatter Plot (extension work)

Warm Up: 5 minutes

The purpose of this warm-up is for students to analyse a scatter plot by noticing things and asking questions. This scatter plot gives students the opportunity to observe a relationship that is not described well with a linear model. As production increases, price tends to decrease. However, there is a lower limit to the price, so the model tends to level out after a certain amount of production is reached.

Instructional Routines

- Notice and Wonder

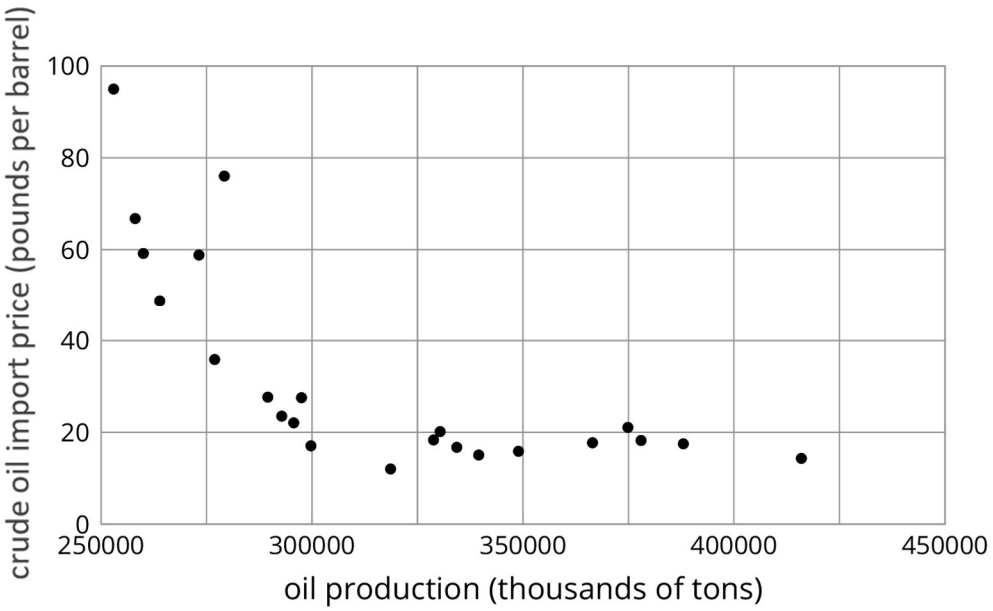
Launch

Display the scatter plot for all to see. Tell students that oil production and price are shown. Production is measured in "thousand tons," so a point on the left side of this graph represents 250 000 thousand tons (or 250 000 000 tons). Price is measured in "pounds per barrel." A barrel of oil is equivalent to 42 gallons. The prices in this chart are for crude oil which would still need to be processed to be made useful for fuel, plastics, or other oil-based materials.

Tell students to signal when they have at least one thing they notice or one thing they wonder. Give students 1 minute of quiet think time followed by whole-class discussion.

Student Task Statement

What do you notice? What do you wonder?



Student Response

Things students may notice:

- The points don't all lie near a line like they have in previous scatter plots.
- As production increases, the price decreases.

Things students may wonder:

- What does "oil production" or "import prices" mean (if the axis labels are unfamiliar)?
- Is it okay to model this relationship using a line?
- Why does it seem to level out at the end?

Activity Synthesis

Ask students to share things they noticed and wondered. Record and display their responses for all to see.

To help focus students on the non-linearity of the association, ask "How is this scatter plot different from ones we have seen so far?" (The data appear to be grouped along a curve and not a straight line.)

Tell students that, even when a linear association seems to be present, it may only fit the data very close to the data that is present. For example, if we cover up the scatter plot to the right of 300 000, there may appear to be a linear association that would fit the data, but it would not apply to the data on the right side of the scatter plot. Additionally, in this particular situation, a linear model does not make sense in the long term, since a linear

model would eventually have a value of zero and then become negative, something that would probably not happen to the price of oil.

7.2 Scatter Plot City

20 minutes

In previous lessons, data was fit with a linear model, but is not appropriate in all cases. In this activity, students individually sort cards of different scatter plots then discuss with a partner. They then re-sort the cards based on linearity and then again by positive or negative association.

Instructional Routines

- Discussion Supports
- Take Turns

Launch

Arrange students in groups of 2, and give each student a set of cards. Give 5 minutes of quiet work time for the first question followed by partner discussion. For the final two questions, ask students to take turns sorting the cards by putting one card into one of the categories, then explaining their reasoning for putting it there. After both partners have agreed on the placement of that card, the other partner may repeat these steps to sort one more card. Repeat this process until all cards are sorted.

Tell the students that there is a third category for the last two parts of the activity: neither. It is possible for data to have absolutely no association and appear completely random which would not be a linear association or a non-linear association. It is also possible for data to not have a strictly positive or negative association if, for example, the scatter plot seems to increase and then decrease.

Representation: Internalise Comprehension. Chunk this task into more manageable parts to differentiate the degree of difficulty or complexity by beginning with fewer cards. For example, give students a subset of the cards to start with and introduce the remaining cards once students have completed their initial set of matches.
Supports accessibility for: Conceptual processing; Organisation Speaking: Discussion Supports. To support students' explanations for sorting the scatter plots in the way they chose, display sentence frames for students to use when they are working with their partner. For example, “I think ____ because _____ .” or “I (agree/disagree) because _____ .”
Design Principle(s): Support sense-making; Optimise output for (explanation)

Student Task Statement

Your teacher will give you a set of cards. Each card shows a scatter plot.

1. Sort the cards into categories and describe each category.

2. Explain the reasoning behind your categories to your partner. Listen to your partner's reasoning for their categories.

3. Sort the cards into two categories: positive associations and negative associations. Compare your sorting with your partner's and discuss any disagreements.

4. Sort the cards into two categories: linear associations and non-linear associations. Compare your sorting with your partner's and discuss any disagreements.

Student Response

1. Answers vary. Sample categories: decimals and whole numbers, positive and negative associations, more and less variable

2. Answers vary.

3. Positive associations: F and G. Negative associations: B, C, D, and E. Neither: A and H

4. Linear associations: C and E. Non-linear associations: A, B, D, F, G, and H

Activity Synthesis

To help clarify the meaning of linear and non-linear associations as well as positive and negative associations, consider asking some of the following questions:

- "How did you originally group the cards?"
- "Which graphs did you sort into the 'positive association' group?" (F and G)
- "Are there any cards that did not fit with either positive or negative association?" (Yes, the points for card A form an arc that first goes up and then down.)
- "Which graphs did you sort into the 'linear association' group?" (C and E)
- "If a scatter plot has no association at all (neither linear nor non-linear), what might its scatter plot look like?" (A lot of random dots scattered all around the plot with no obvious trend.)

7.3 Clustering

10 minutes

In addition to the type of model appropriate for the data represented in a scatter plot and the positive or negative association, there are other patterns that can be worth noticing before performing a more in-depth analysis. In this activity, students compare scatter plots that have more than one pattern embedded in them. In the discussion following the activity, this feature is identified as *clustering*.

As students compare the scatter plots in this activity, monitor for students who notice clustering of the data and name it in their own words.

Instructional Routines

- Collect and Display
- Think Pair Share

Launch

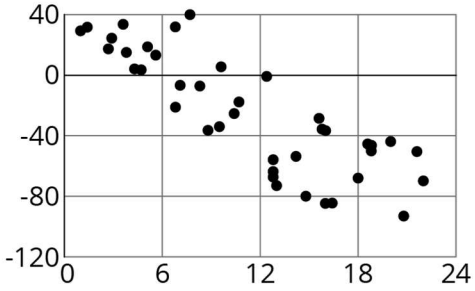
Keep students in groups of 2. Allow 2 minutes quiet think time followed by partner and whole-class discussion.

Engagement: Develop Effort and Persistence. Encourage and support opportunities for peer interactions. Display sentence frames to support student conversation such as: "_____ and _____ are different because…", "One thing that is the same is…", and "What do _____ and _____ have in common?"
Supports accessibility for: Language; Social-emotional skills Conversing, Representing, Writing: Collect and Display. As students discuss how the scatter plots are alike, and how they are different, capture the vocabulary and phrases students use to describe the patterns they notice. Listen for students who refer to clustering, or multiple patterns in one graph. Record their language and any relevant sketches on a visual display that can be referenced in future discussions. This will help students produce and make sense of the language needed to communicate about the relationship between quantities represented graphically.
Design Principle(s): Support sense-making; Maximise meta-awareness

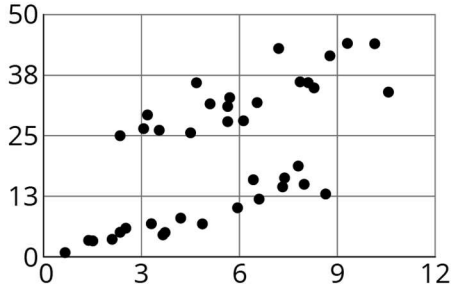
Student Task Statement

How are these scatter plots alike? How are they different?

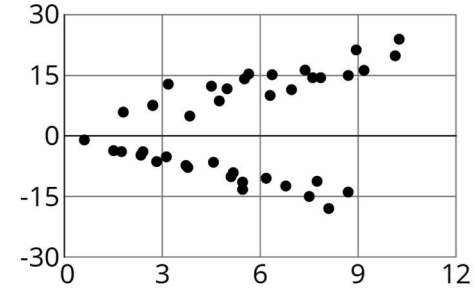
A



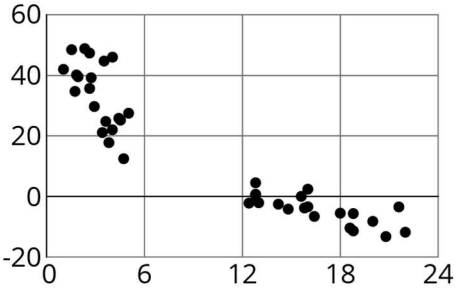
B



C



D



Student Response

Answers vary. Sample responses:

- Plots A and D both have negative associations, but D has a gap in the data.
- Plots B and C both have 2 groups of data that would have different lines of fit. In B, both lines of fit would have about the same gradient, but in C, one gradient is positive and the other is negative.
- Plots A, C, and D each have data in quadrants I and IV, but plot B only has data in quadrant I.

Activity Synthesis

The purpose of this discussion is for students to understand what it means for data to appear in clusters and recognise when they might appear.

Select previously identified students to share their observations about the scatter plots. Tell students that when data seems to have more than one pattern, it is called *clustering*. Clustering of the data like in graph B, C, and D can reveal hidden patterns. Usually, clustering means there are subgroups within our data that may represent different trends.

For example, in plot B, the data may represent body measurements of a certain species of bird. Although the data originally came from a group that made sense (a single species), there appear to be subgroups that have a large influence on the data as well. The lower half of the data may represent females and the upper half may represent males, so we can see that there are different patterns within the different subgroups.

When clustering is present, it may be helpful to investigate the cause of the separation and analyse the data within the subgroups rather than as a whole.

Lesson Synthesis

To help summarise the lesson, ask:

- "In your own words, what is a non-linear association?" (The points in a scatter plot do not lie along a straight line.)
- "What does a non-linear association mean?" (There is not a constant increase or decrease of one variable based on the other.)
- "In your own words, what are clusters in data?" (The points in a scatter plot are clumped together in different groups.)
- "What do clusters usually mean?" (There may be multiple patterns present within the data. Perhaps there are subgroups that show different patterns.)

7.4 Make Your Own Scatter Plot

Cool Down: 5 minutes

Students show their understanding of positive and negative association, linear and non-linear association, and clustering of data by drawing scatter plots with these conditions.

Launch

Advise students to use around 20 data points in their graph. This is enough to show the intention of the graph without being tedious.

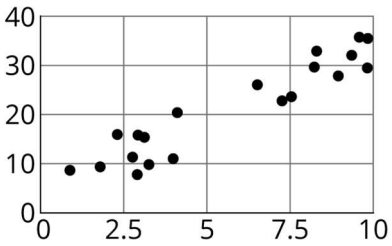
Student Task Statement

1. Draw a scatter plot that shows a positive linear association and clustering.

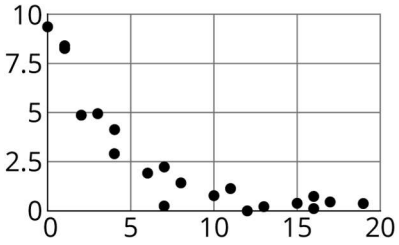
2. Draw a scatter plot that shows a negative non-linear association and no clustering.

Student Response

1. Answers vary. Sample response:

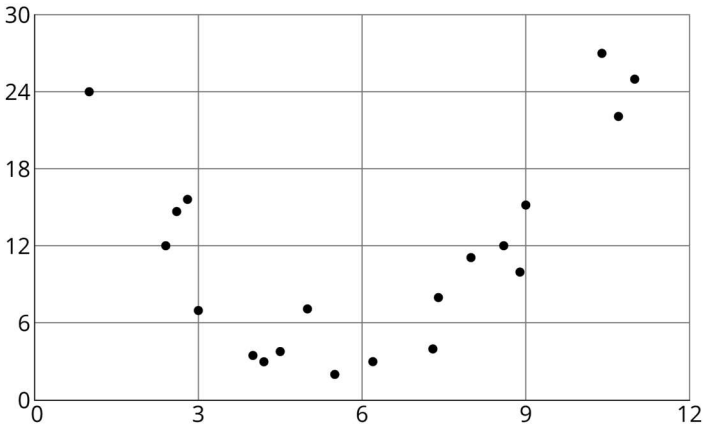


2. Answers vary. Sample response:



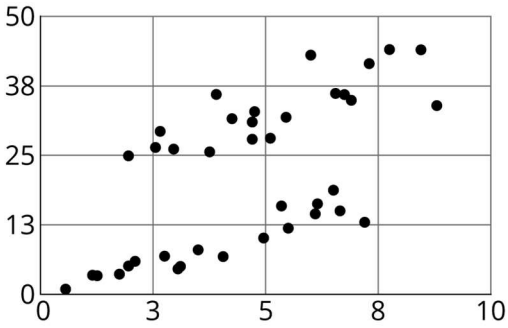
Student Lesson Summary

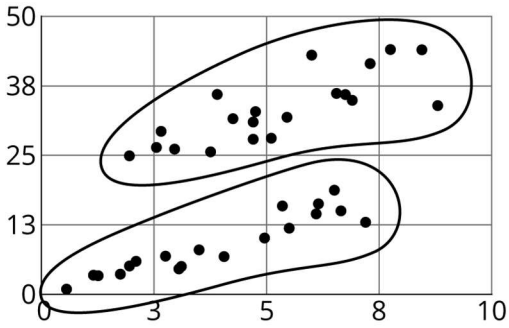
Sometimes a scatter plot shows an association that is *not* linear:



We call such an association a *non-linear association*. Later you will study functions that can be models for non-linear associations.

Sometimes in a scatter plot we can see separate groups of points.



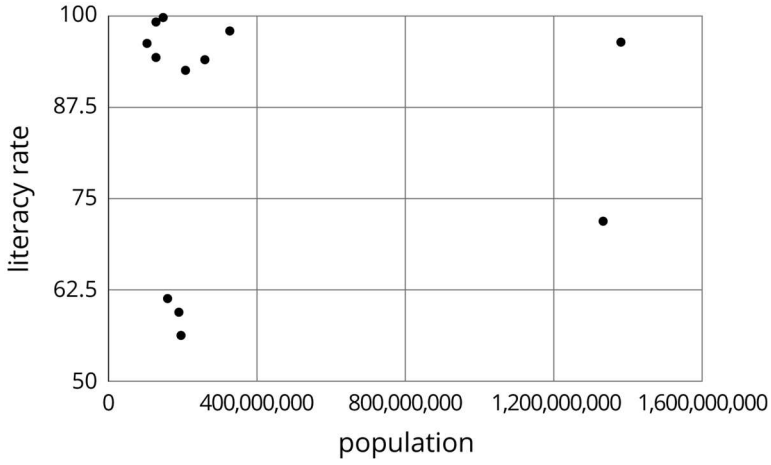


We call these groups *clusters*.

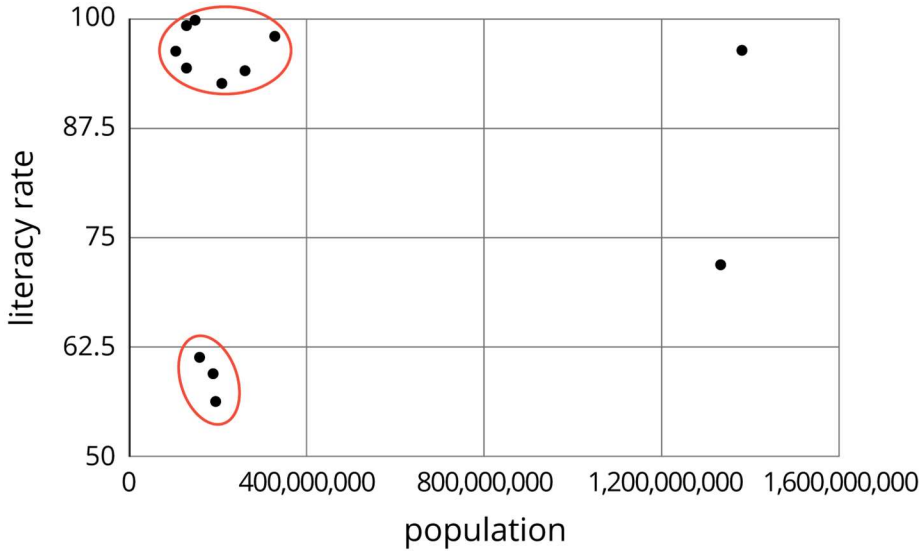
Lesson 7 Practice Problems

Problem 1 Statement

Literacy rate and population for the 12 countries with more than 100 million people are shown in the scatter plot. Circle any clusters in the data.

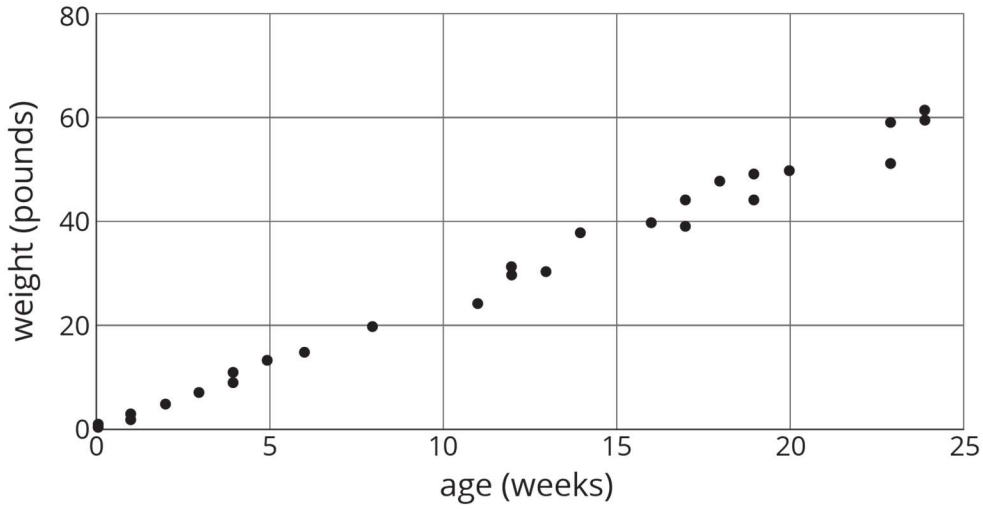


Solution



Problem 2 Statement

Here is a scatter plot:



Select **all** the following that describe the association in the scatter plot:

a. Linear association

b. Non-linear association

c. Positive association

d. Negative association

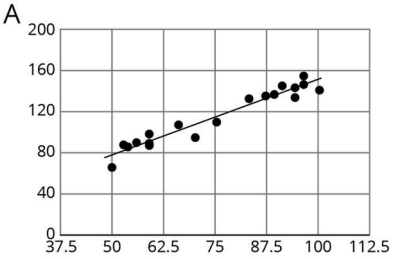
e. No association

Solution ["A", "C"]

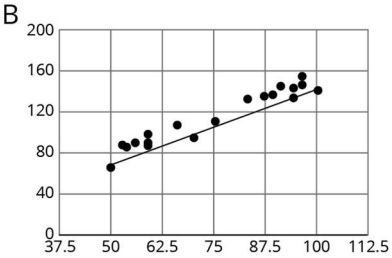
Problem 3 Statement

For the same data, two different models are graphed. Which model more closely matches the data? Explain your reasoning.

A



B



Solution

Model A more closely matches the data. In model B, most of the points are above the line in the graph. In model A, the points are more evenly arranged around the line.

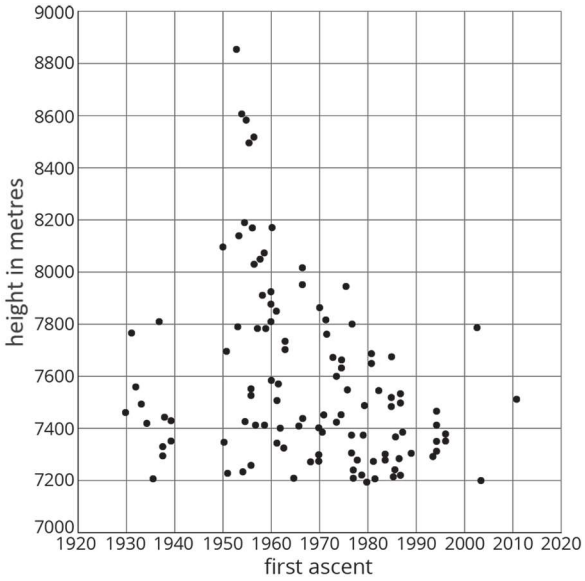
Problem 4 Statement

Here is a scatter plot of data for some of the tallest mountains on Earth.

The heights in metres and year of first recorded ascent is shown. Mount Everest is the tallest mountain in this set of data.

a. Estimate the height of Mount Everest.

b. Estimate the year of the first recorded ascent of Mount Everest.



Solution

a. Approximately 8 848 metres (The vertical coordinate of the data point with the greatest vertical value is closer to 8 800 than 9 000.)

b. Approximately 1953 (The horizontal coordinate of the same point is slightly to the right of 1950.)

Problem 5 Statement

A cone has a volume V, radius r, and a height of 12 cm.

a. A cone has the same height and $\frac{1}{3}$ of the radius of the original cone. Write an expression for its volume.

b. A cone has the same height and 3 times the radius of the original cone. Write an expression for its volume.

Solution

a. $\frac{V}{9}$

b. $9V$