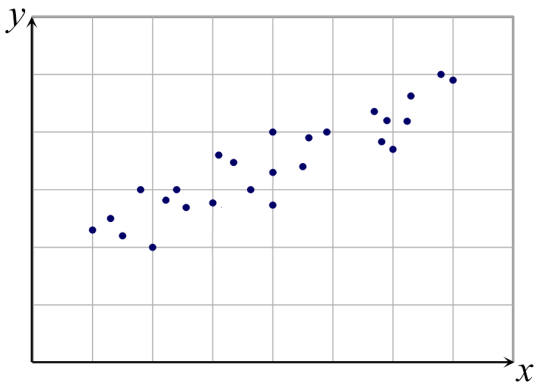
Scatter Plot Features and Patterns

When studying data in two variables, we are usually interested in knowing whether there is any association between them; in other words, whether the two things we are studying are connected in any way.

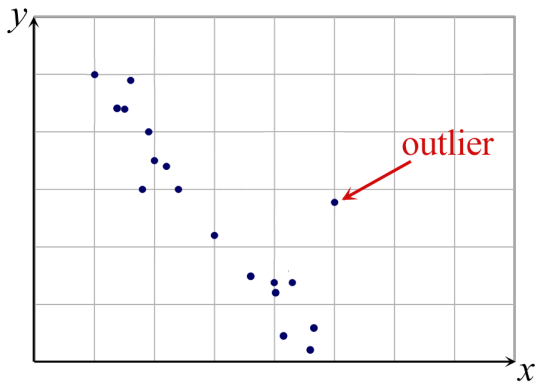
An association will be seen visually in a scatter plot when the points lie in a visible pattern. The data can also have outliers and clusters.

A positive linear association



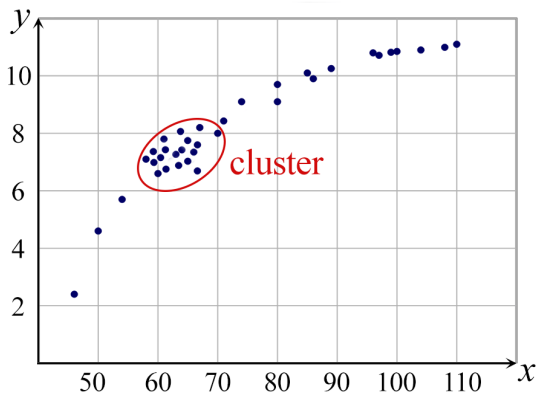
The pattern of the points is as if on a line (linear). As the *x* values increase, the *y*-values increase also, so the association is **positive**.

A negative linear association



The points lie in a linear pattern. As the *x* values increase, the *y*-values *decrease*, so the association is **negative**. One point lies far from the others; it is an **outlier**.

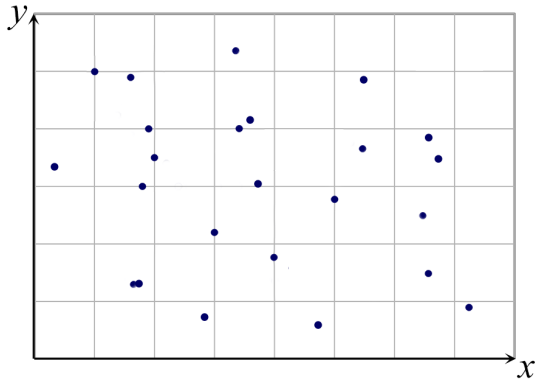
A positive nonlinear association



Here as the *x*-values increase, the *y*-values increase also, so the association is positive. But the pattern of dots follows a curve, not a straight line, so we say that the association is **nonlinear**.

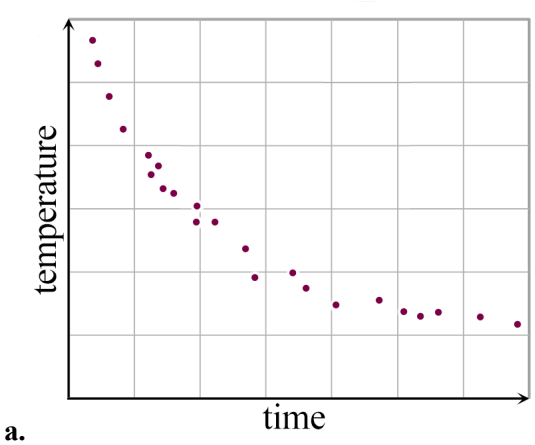
Additionally, we see a **cluster** — many points close together in a small area, around the *x* values of 58 to 70 and *y*-values of 7.5 to 8.3.

No association

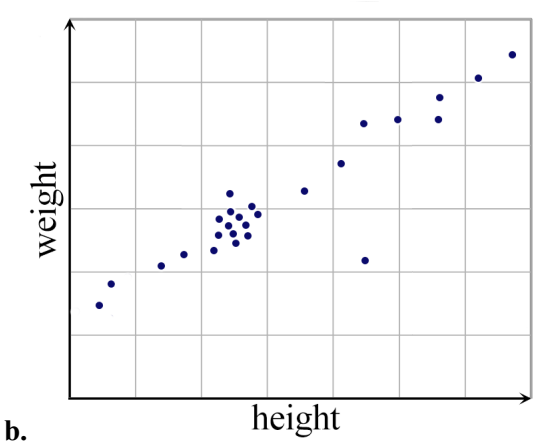


Here, there is no association between the variables. The dots are scattered and do not lie in any visible pattern.

1. Describe the patterns and features you see in each scatter plot.

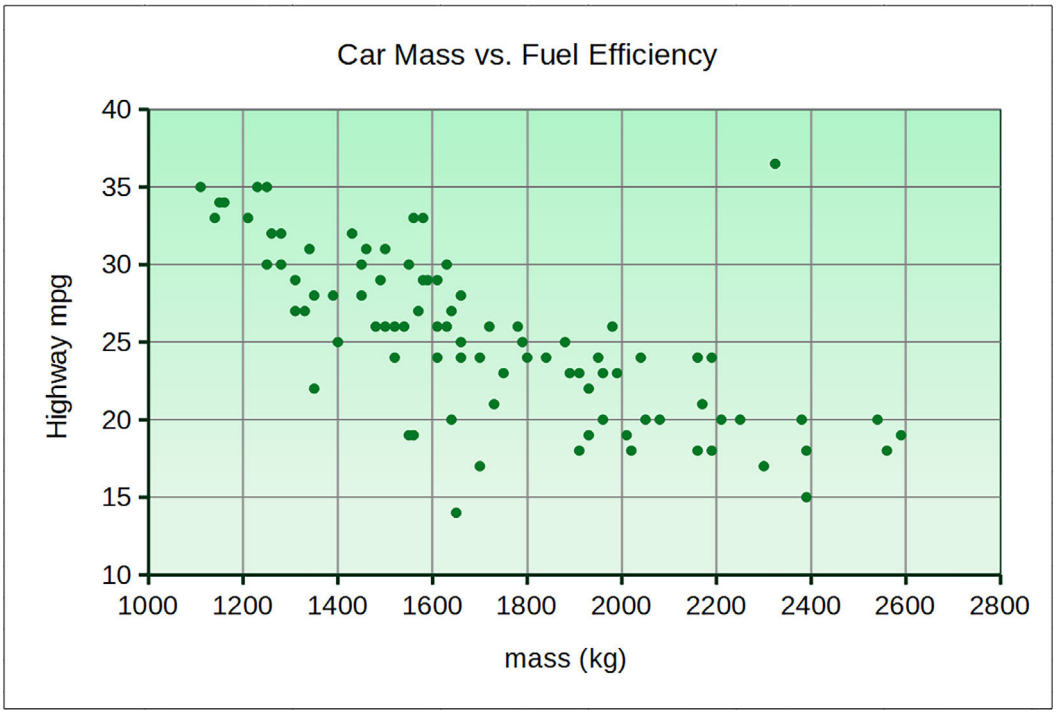


a.



b.

2. This graph shows the mass and fuel efficiency (in miles per gallon, mpg, for highway driving) of various cars.



a. Describe the general pattern and any special features of the plot.
Do heavier cars get better or worse gas mileage, in general, than lighter cars?

b. Find the heaviest car with 30 mpg gas mileage. How much does it weigh?

c. List the mass and the gas mileage of the car that is so different from the others that it may be even an error in the data, such as a typo.

d. Aside from the outlier, what is the best fuel efficiency for cars weighing from 2000 to 2400 kg?

3. The data shows the height and the shoe size of a group of individuals. The shoe size was measured in centimeters, to the nearest half centimeter, instead of using shoe size, because shoe sizes vary quite a bit between different brands and countries.

   Make a scatter plot. Choose the scaling for both axes wisely. You don't want the data points scrunched up in one area of the graph and most of it being empty. The data points should "fill" the graph area, the best possible, so that the scatter plot features will be visible well.

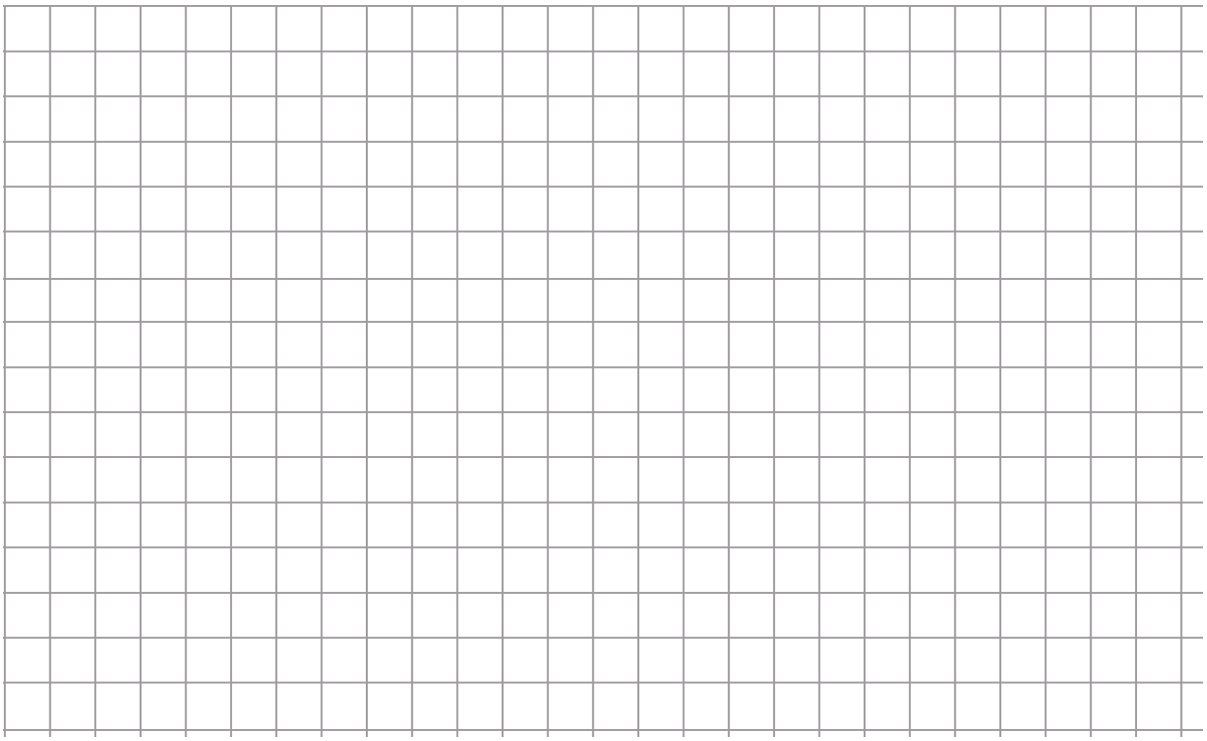
   In this case, you don't want to start the scaling from 0, for either variable. Check the minimum and the maximum values (for both variables), and use those to guide you in regards to scaling.

Height (cm)	Shoe size (cm)
155	23
158	23.5
160	22.5
160	24.5
162	23.5
162	24
164	24

Height (cm)	Shoe size (cm)
165	23
165	24.5
167	24
167	25
167	25.5
168	26
168	26

Height (cm)	Shoe size (cm)
170	27
171	26
173	26.5
175	26
177	25.5
178	28
179	27.5

Height (cm)	Shoe size (cm)
180	28
182	27
185	28
186	29
186	29
188	29.5



   Lastly, describe the basic pattern and possible special features of the scatter plot.