

# Describing Scatterplots

by Sophia



## WHAT'S COVERED

In this tutorial, you're going to learn about describing scatterplots. Our discussion breaks down as follows:

1. Describing Scatterplots
2. Form
3. Direction
3. Strength

## 1. Describing Scatterplots

When talking about univariate data, or one-variable data, you would discuss the shape, center, and spread of a distribution when making histograms and dot plots,

On a scatterplot, it's a bit difficult to talk about the shape. Regarding center and spread, it's all very confusing; perhaps the QB salary is very spread out, and the total salary is maybe not so spread out. This would make it hard to talk about the spread.

Instead, you're actually going to describe:

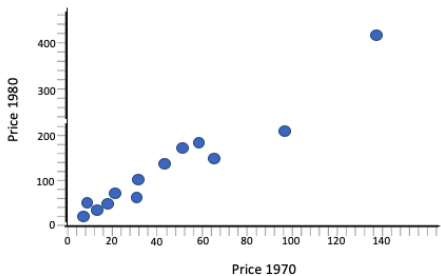
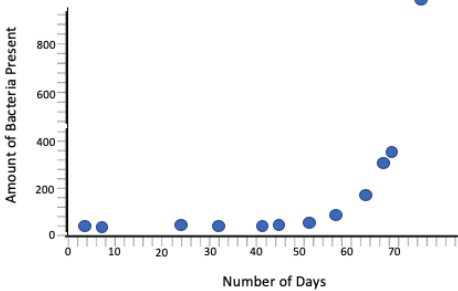
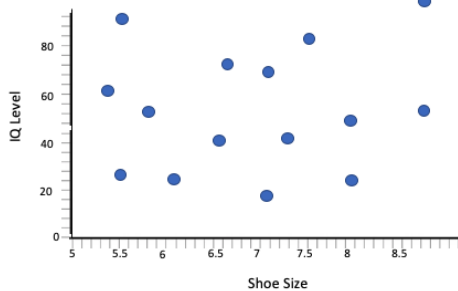
- Form
- Direction
- Strength

---

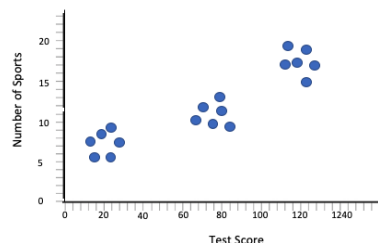
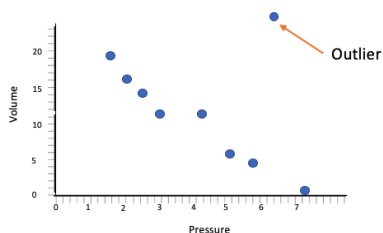
## 2. Form

In the **form**, we look for a pattern. Is the pattern linear, or do the data show a curve? Do they start low, then peak, then end low? Or do they start low and end high? How do they curve or do they rise quickly and then tail off? There's a lot to look at.

When discussing form, you will most likely describe a scatterplot as linear or non-linear.

Forms of a Scatterplot	
<p><b>Linear:</b> The scatterplot is approximating a line.</p>	
<p><b>Non-Linear:</b> The data points follow a curve.</p>	
<p><b>No Association:</b> Data points resemble a cloud and there is not clear pattern.</p>	

In addition, it is important to consider outliers or clusters when looking at the form. If the scatterplot was essentially linear but had one outlier, we would want to note that. Also, if the data created clusters throughout the scatterplot, this will be important to keep track of.



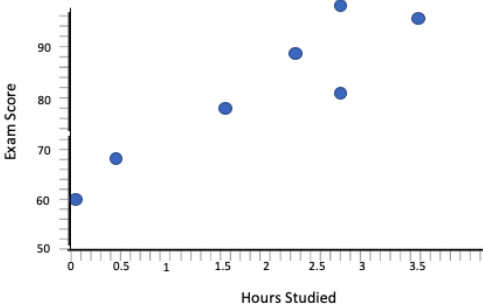
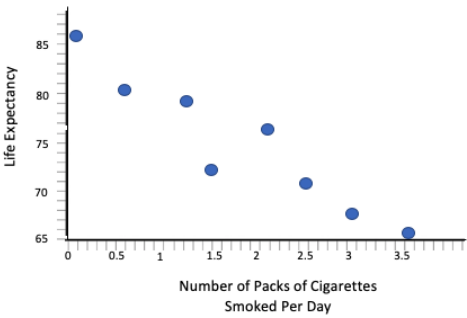
## TERM TO KNOW

### Form

The overall shape of the data points. The form may be linear or nonlinear, or there may not be any form at all to the points if they form a "cloud."

### 3. Direction

The **direction** refers to how the y-axis variable responds as you move to the right on the x-axis variable. There are two main directions that a scatterplot can have are positive and negative.

Direction of a Scatterplot																			
<b>Positive:</b> The variables both increase or both decrease	 <table border="1"><caption>Data for Positive Scatterplot</caption><thead><tr><th>Hours Studied</th><th>Exam Score</th></tr></thead><tbody><tr><td>0.2</td><td>60</td></tr><tr><td>0.5</td><td>68</td></tr><tr><td>1.5</td><td>78</td></tr><tr><td>2.2</td><td>88</td></tr><tr><td>2.5</td><td>82</td></tr><tr><td>2.8</td><td>92</td></tr><tr><td>3.5</td><td>90</td></tr></tbody></table>	Hours Studied	Exam Score	0.2	60	0.5	68	1.5	78	2.2	88	2.5	82	2.8	92	3.5	90		
Hours Studied	Exam Score																		
0.2	60																		
0.5	68																		
1.5	78																		
2.2	88																		
2.5	82																		
2.8	92																		
3.5	90																		
<b>Negative:</b> The variables go in opposite directions (one variable increases while the other variable decreases)	 <table border="1"><caption>Data for Negative Scatterplot</caption><thead><tr><th>Number of Packs of Cigarettes Smoked Per Day</th><th>Life Expectancy</th></tr></thead><tbody><tr><td>0.2</td><td>85</td></tr><tr><td>0.5</td><td>80</td></tr><tr><td>1.2</td><td>79</td></tr><tr><td>1.5</td><td>72</td></tr><tr><td>2.2</td><td>76</td></tr><tr><td>2.5</td><td>71</td></tr><tr><td>3.0</td><td>68</td></tr><tr><td>3.5</td><td>65</td></tr></tbody></table>	Number of Packs of Cigarettes Smoked Per Day	Life Expectancy	0.2	85	0.5	80	1.2	79	1.5	72	2.2	76	2.5	71	3.0	68	3.5	65
Number of Packs of Cigarettes Smoked Per Day	Life Expectancy																		
0.2	85																		
0.5	80																		
1.2	79																		
1.5	72																		
2.2	76																		
2.5	71																		
3.0	68																		
3.5	65																		



#### TERM TO KNOW

##### Direction

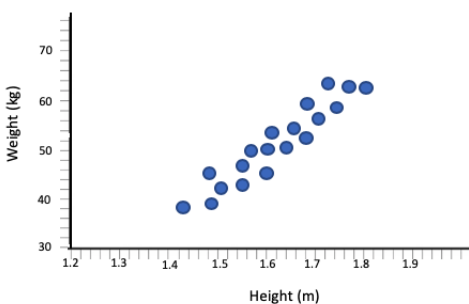
The way one variable responds to an increase in the other. With a negative association, an increase in one variable is associated with a decrease in the other, whereas with a positive association, an increase in one variable is associated with an increase in the other.

### 3. Strength

The **strength** is how closely the two variables are associated with some line or curve. How well do the points follow that indicated form? How well do these points stack up on a line? The strength can be described as strong, moderate, or weak.

#### Strengths of a Scatterplot

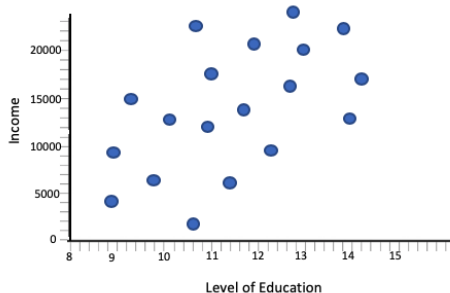
**Strong:** The scatterplot would most resemble the form. The data points are clustered around either a line or a curve.



**Moderate:** The data points are less clustered in a line or curve, however, the direction is still clear.



**Weak:** The data points are much more spread out and the direction may be less clear.



#### THINK ABOUT IT

Imagine the oval that you could put over the scatterplots. A strong association would have a very long, thin oval over it. The moderate association would have kind of a wider oval over it, but it would still be longer than it is wide. Over the weak association, the oval is almost more like a circle.



#### HINT

The idea is, if you can encase the points in an oval, the stronger associations will have a longer, thinner oval.



#### TERM TO KNOW

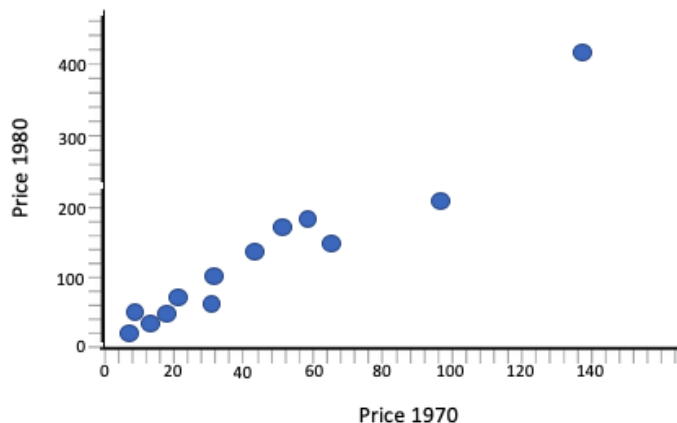
##### Strength

The closeness of the points to the indicated form. Points that are strongly linear will all fall on or near a straight line.



#### TRY IT

This is the 1970 and 1980 price of different seafood in cents per pound. How would you describe the form, direction, and strength of this scatterplot?

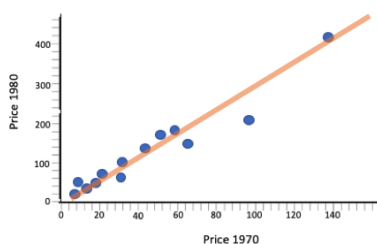


## Form, Direction, and Strength of Scatterplot

(1970 vs 1980 Seafood Prices)

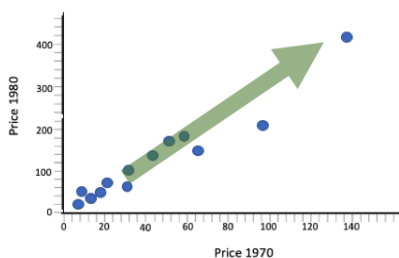
### Form: Linear

The form is fairly linear. One point is a little bit low for the line that we would look at for the rest of the data points. It also appears to have an outlier on the high side.



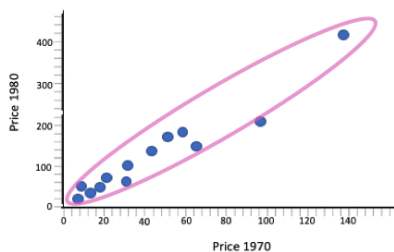
### Direction: Positive

The direction is positive, which means that as the 1970 price increases, so does the 1980 price. That's not surprising, because you would expect that the ones that are less expensive in 1970 would also be less expensive in 1980, and the ones that were more expensive in '70 would be more expensive in '80.



### Strength: Strong

The strength is very strong because it's fairly predictable what is going to happen with these prices, based on the fact that they're very close to a line.





## SUMMARY

To describe scatterplots that look at one variable data, we look at shape, center, and spread. When we look at two variable data scatterplots, we analyze form, direction, and strength. Regarding form, are they linear or nonlinear? Are there unusual features, gaps, clusters, or outliers? Regarding strength, how well do they follow that form? And lastly, we analyze the direction of the association: what happens as the x-axis variable increases? Does the y-axis variable go up, down, or does it stay the same? Or, is there really no association at all?

Good luck!

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR [TERMS OF USE](#).



## TERMS TO KNOW

### Direction

The way one variable responds to an increase in the other. With a negative association, an increase in one variable is associated with a decrease in the other, whereas with a positive association, an increase in one variable is associated with an increase in the other.

### Form

The overall shape of the data points. The form may be linear or nonlinear, or there may not be any form at all to the points, if they form a "cloud."

### Strength

The closeness of the points to the indicated form. Points that are strongly linear will all fall on or near a straight line.