










Scatter Plots

Students investigate scatter plots as a method of visualizing the relationship between two quantitative variables.

Prerequisites	Displaying Categorical Data																		
Relevant Standards <div>OK K12CS CSTA CC-Math NGSS</div>	Select one or more standards from the menu on the left (⌘-click on Mac, Ctrl-click elsewhere).																		
Lesson Goals	Students will be able to... <ul style="list-style-type: none">consider explanatory and response roles of variablesmake scatter plots by hand, given a list of (x,y) pairsmake scatter plots using Pyretidentify a possible linear relationship by looking at a point cloudConsider unusual observations in a scatter plot																		
Student-facing Lesson Goals	<ul style="list-style-type: none">Let’s use Pyret to create scatter plots of data.																		
Materials	<ul style="list-style-type: none">Lesson Slides (Google Slides)Computer for each student (or pair), with access to the internetStudent workbook, and something to write with																		
Preparation	<ul style="list-style-type: none">Make sure all materials have been gatheredDecide how students will be grouped in pairsAll students should log into CPO and open the "Animals Starter File" they saved from the prior lesson. If they don’t have the file, they can open a new one																		
Supplemental Resources																			
Language Table	<table><tr><th>Types</th><th>Functions</th><th>Values</th></tr><tr><td>Number</td><td>num-sqrt, num-sqr, mean, median, modes</td><td>4, -1.2, 2/3</td></tr><tr><td>String</td><td>string-repeat, string-contains</td><td>"hello", "91"</td></tr><tr><td>Boolean</td><td>==, <, <=, >=, string-equal</td><td>true, false</td></tr><tr><td>Image</td><td>triangle, circle, star, rectangle, ellipse, square, text, overlay, bar-chart, pie-chart, bar-chart-summarized, pie-chart-summarized, histogram</td><td>  </td></tr><tr><td>Table</td><td>count, .row-n, .order-by, .filter, .build-column</td><td></td></tr></table>	Types	Functions	Values	Number	num-sqrt, num-sqr, mean, median, modes	4, -1.2, 2/3	String	string-repeat, string-contains	"hello", "91"	Boolean	==, <, <=, >=, string-equal	true, false	Image	triangle, circle, star, rectangle, ellipse, square, text, overlay, bar-chart, pie-chart, bar-chart-summarized, pie-chart-summarized, histogram	  	Table	count, .row-n, .order-by, .filter, .build-column	
Types	Functions	Values																	
Number	num-sqrt, num-sqr, mean, median, modes	4, -1.2, 2/3																	
String	string-repeat, string-contains	"hello", "91"																	
Boolean	==, <, <=, >=, string-equal	true, false																	
Image	triangle, circle, star, rectangle, ellipse, square, text, overlay, bar-chart, pie-chart, bar-chart-summarized, pie-chart-summarized, histogram	  																	
Table	count, .row-n, .order-by, .filter, .build-column																		

Glossary

explanatory variable :: the variable in a relationship that is presumed to impact the other variable

response variable :: the variable in a relationship that is presumed to be affected by the other variable

scatter plot :: a display of the relationship between two quantitative variables, graphing each explanatory value on the x axis and the accompanying response on the y axis

Relationships *Between Columns*

15 minutes

Overview

Students are finally introduced to *Relate Questions*, which ask about the relationship between one quantitative column and another.

Launch

Can animals' weights help explain why some are adopted quickly while others take a long time? What other factors explain why one pet gets adopted right away, and others wait months?

Theory 1: Smaller animals get adopted faster because they're easier to care for.

How could we test that theory? Bar and pie charts are great for showing us frequencies or percentages in a categorical column. Histograms and box plots are great for showing us the shape, center, and spread of a single quantitative column. But none of these displays will help us see connections between **two** quantitative columns.

Investigate

- Take a few minutes to look through the whole dataset, and see if you agree with Theory 1.
- Could any of our visualizations or summaries provide evidence for or against the theory?
- Write down your hypothesis on [\(Dis\)Proving a Claim \(Page 71\)](#), as well as a theory about how we could use this dataset to see if you're right.

Synthesize

We've got a lot of tools in our toolkit that help us think about an entire *column* of a dataset:

- We have ways to find measures of center and spread for a given quantitative column.
- We have visualizations that let us see the shape of values in a quantitative column.
- We have visualizations that let us see frequencies or percentages in a categorical column.

What columns is this question asking about?

Making Scatter Plots

20 minutes

Overview

Students are introduced to *scatter plots*, which are visualizations tailored to *Relate Questions* about quantitative variables. They learn how to construct scatter plots by hand, and in Pyret.

Launch

This question is asking about *two columns* in our dataset. Specifically, it's asking **if there is a relationship** between `pounds` and `weeks`.

Before we can draw a *scatter plot*, we have to make an important decision: which variable is *explanatory* and which is *response*? In this case, are we suspecting that an animal's weight can explain how long it takes to be adopted, or that how long it takes to be adopted can explain how much an animal weighs?

The first of these makes sense, and reflects our suspicion that weight plays a role in adoption time. The convention is to use the horizontal axis for our explanatory variable and the vertical axis for the response. Thus, `pounds` will be `x` and `weeks` will be `y`.

Investigate

We will produce our scatter plot by graphing each animal's `pounds` and `weeks` values as a point on the `x` and `y` axes.

Complete [Creating a Scatter Plot \(Page 72\)](#) in your Student Workbook.

Teaching Tip

Divide the full table up into sub-lists, and have a few students plot 3-4 animals on the board. This can be done collaboratively, resulting in a whole-class scatterplot!

- Open your “Animals Starter File”. (If you do not have this file, or if something has happened to it, you can always make a [new copy](#).)
- Make a scatter plot that displays the relationship between weight and adoption time.
- Are there any patterns or trends that you see here?
- Try making a few other scatter plots, looking for relationships between other columns in the `animals-table`.

Synthesize

Have students share their observations. What trends do they see? Are there any points that seem unusual? Why?

Looking for Trends

20 minutes

Overview

Students are asked to identify patterns in their scatter plots. This activity builds towards the idea of *linear associations*, but does not go into depth (as the following lesson does).

Launch

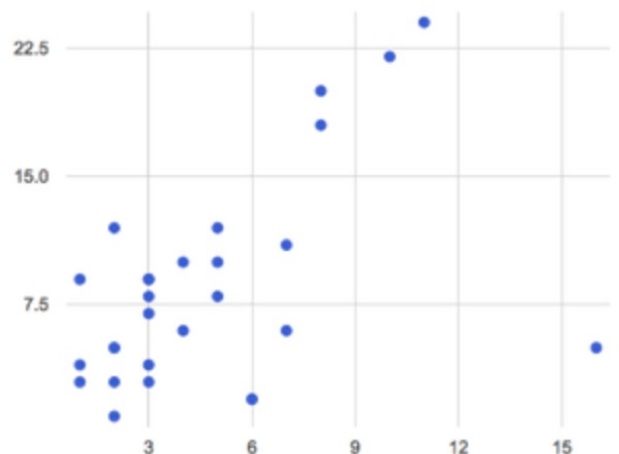
- Can you see a “cloud” around which the points are clustered?
- Does the number of weeks to adoption seem to go up or down as the weight increases?
- Are there any points that “stray from the pack”? Which ones?

Teaching Tip

Project the scatter plot at the front of the room, and have students come up to the plot to point out their patterns.

A straight-line pattern in the cloud of scatter plot points suggests that there is a linear relationship between two columns. If we can pinpoint a line around which the points cluster (as we'll do in a future lesson), it would be useful for making predictions. For example, our line might predict how many `weeks` a new dog would wait to be adopted, if the dog weighs `68 pounds`.

Do any data points seem unusually far away from the main cloud of points? Which animals are those? These points are called **unusual observations**. Unusual observations in a scatter plot are like outliers in a histogram, but more complicated because it's the *combination* of *x* and *y* values that makes them stand apart from the rest of the cloud.



Unusual observations are *always* worth thinking about

- Sometimes they're just random. Felix seems to have been adopted quickly, considering how much he weighs. Maybe he just met the right family early, or maybe we find out he lives nearby, got lost and his family came to get him. In that case, we might need to do some deep thinking about whether or not it's appropriate to remove him from our dataset.
- Sometimes they can give you a deeper insight into your data. Maybe Felix is a special, popular (and heavy!) breed of cat, and we discover that our dataset is missing an important column for breed!

and we discover that our dataset is missing an important column for breed.

- Sometimes unusual observations are the points we are looking for! What if we wanted to know which restaurants are a good value, and which are rip-offs? We could make a scatter plot of restaurant reviews vs. prices, and look for an observation that's high above the rest of the points. That would be a restaurant whose reviews are *unusually good* for the price. An observation way below the cloud would be a really bad deal.

Investigate

For practice, try making scatter plots for each of the following relationships, always expressed as “response variable vs explanatory variable”. If you see any **unusual observations**, try to explain them!

- The `pounds` of an animal vs its `age`
- The number of `weeks` for an animal to be adopted vs its number of `legs`
- The number of `legs` vs the `age` of an animal.
- Do you see a linear (straight-line) relationship in any of these, evidenced by a cloud of points that's clearly rising or falling from left to right? Are there any unusual observations?

Synthesize

Debrief, showing the plots on the board. Make sure students see plots for which there is no relationship, like the last one!

Theory 2: Younger animals get adopted faster because they are cuter.
--

It might be tempting to go straight into making a scatter plot to explore how weeks to adoption may be affected by age. But different animals have very different lifespans! A 5-year-old tarantula is still really young, while a 5-year-old rabbit is fully grown. With differences like this, it doesn't make sense to put them all on the same scatter plot. By mixing them together, we may be *hiding* a real relationship, or creating the illusion of a relationship that isn't really there! So it's probably best to make several displays, one for each species.