Grouped Samples

Students learn about grouped samples, and practice creating them from the Animals Dataset. In the process, they practice using the Design Recipe to create filter functions, and come up with questions they wish to explore.

Prerequisites	Defining Table Functions
Relevant Standards OK K12CS CSTA NGSS CC-Math	*Select one or more standards from the menu on the left (⌘-click on Mac, Ctrl-click elsewhere).*
Lesson Goals	Students will be able to... • Make grouped samples from a population
Student-facing Lesson Goals	• Let's combine what we know about sampling and filtering with creating displays.
Materials	• Computer for each student (or pair), with access to the internet • Student workbook, and something to write with
Preparation	• Lesson slides (Google Slides) • Make sure all materials have been gathered • Decide how students will be grouped in pairs • All 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	*(see table below)*

Types	Functions	Values
Number	`num-sqrt, num-sqr`	`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`	◉△◇
Table	`count, .row-n, .order-by, .filter, .build-column`	

Glossary

grouped sample :: a non-random subset of individuals chosen from a larger set, where the individuals belong to a specific group

---

# Problems with a Single Population   10 minutes

Overview

This activity is all about *grouped samples*: Students make a bunch of subsets from the Animals Dataset, and see how each subset might answer the same question differently.

Launch

When looking at a scatter plot of our animals, it looks like the amount an animal weighs may have something to do with how long it takes to be adopted.

But if we label the dots by animal (see the image on the right), we notice *every data point* after 25 pounds belongs to a dog from the shelter!
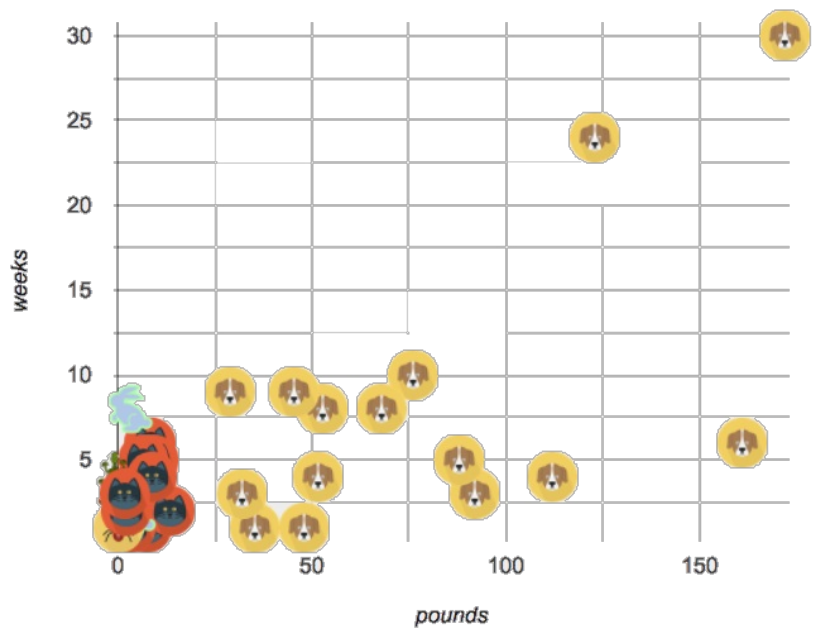
Investigate

Divide the class into groups of 3-4, with one student identified as the "reporter".

- Looking at this scatterplot, does it make sense to analyze all the animals together? Why or why not?

- Are there some questions where it would be important to break up the population into species-specific populations? What are they?

- Are there some questions where it would be important to keep the whole population together? What are they?

Synthesize

Have the reporters share their findings with the class.

Imagine that you've been handed a dataset from a country where half the people are wealthy and have access to amazing medical care, and the other half are poor and have no healthcare. If we took a random sample of the population as a whole, we might think that they are generally middle-income and have average health. But if we ask the same question about the two groups *separately* , we would discover inequality hiding in plain sight!

Grouped Samples   20 minutes

Launch

Ultimately, it might make more sense to certain questions about "just the cats" or "just the dogs". Averaging every animal together will give us an answer, but it may not be a *useful* answer.

> Sometimes important facts about samples get *lost* if we mix them with the rest of the population!

Data Scientists make *grouped samples* of datasets, breaking them up into sub-groups that may be helpful in their analysis.

Investigate

A "kitten" is an animal who *is a cat* and who *is young* . How would you make a subset of just kittens?

- Turn to Grouped Samples from the Animals Dataset (Page 41), and see what code will compute whether or not an animal is a kitten.

- Can you fill in the code for the other subsets?

- When you're done, type these definitions into the Definitions Area

We already know how to define values, and how to filter a dataset. So let's put those skills together to define one of our subsets:

```
dogs  = animals-table.filter(is-dog)
```

- Define the other subsets, and click "Run".
- Make a pie chart showing the species in the `young` subset, by typing `pie-chart(young, "species")`.
- Make pie charts for every grouped sample. Which one is the most representative of the whole population? Why?

Synthesize
Debrief with students. Thoughtful question: how could we filter *and* sort a table? How can we combine methods?

Displaying Samples   20 minutes

Overview
Students revisit the data display activity, now using the samples they created.

Launch
Making grouped and random samples is a powerful skill to have, which allows us to dig deeper than just making charts or asking questions about a whole dataset. Now that we know how to make subsets, we can make much more sophisticated displays!

Investigate
Complete Displaying Data (Page 42), using what you've learned about samples to make more sophisticated data displays.

Synthesize
Were any of the students' displays interesting or surprising? Given a novel question, can students identify what helper functions they would need to write?