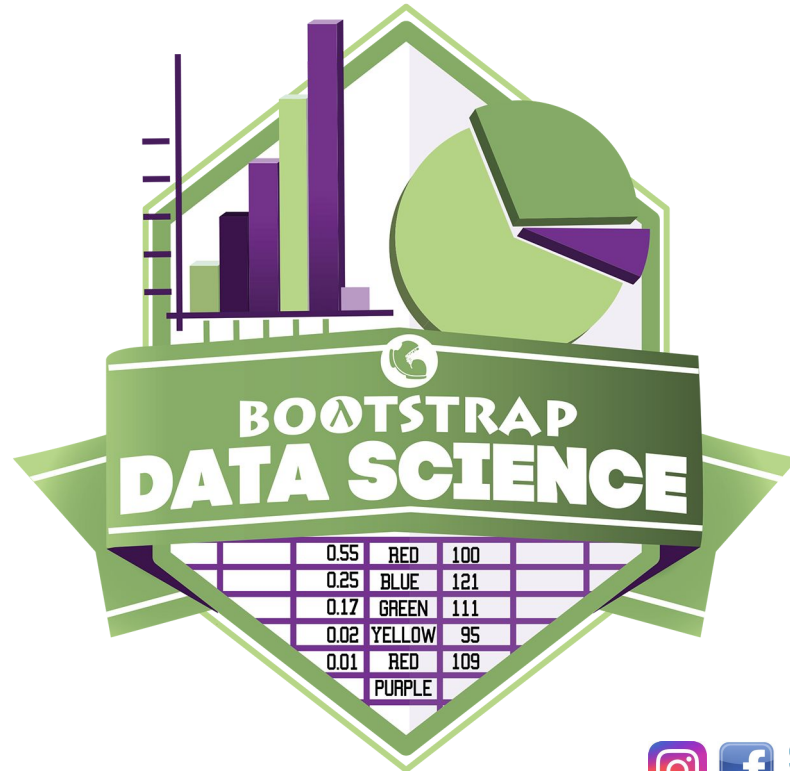


Grouped Samples

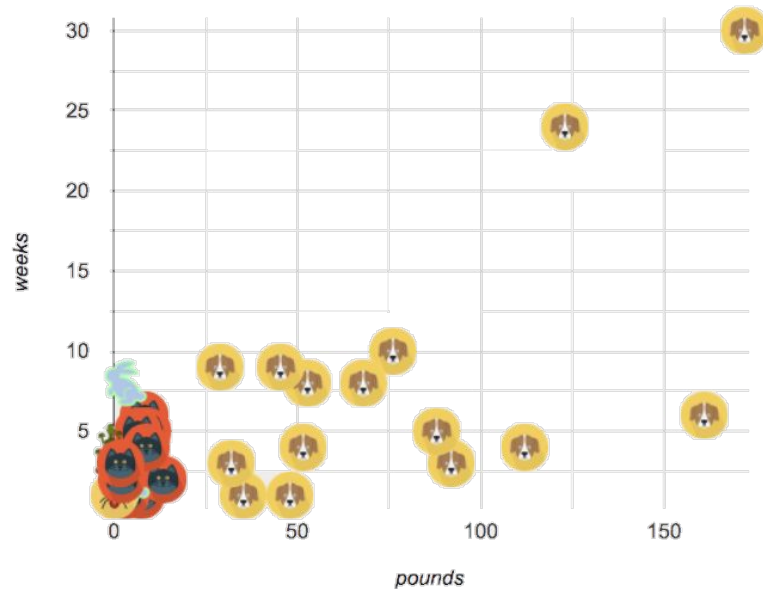




Problems with a Single Population

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, we notice **every datapoint after 25 pounds belongs to a dog** at the shelter!

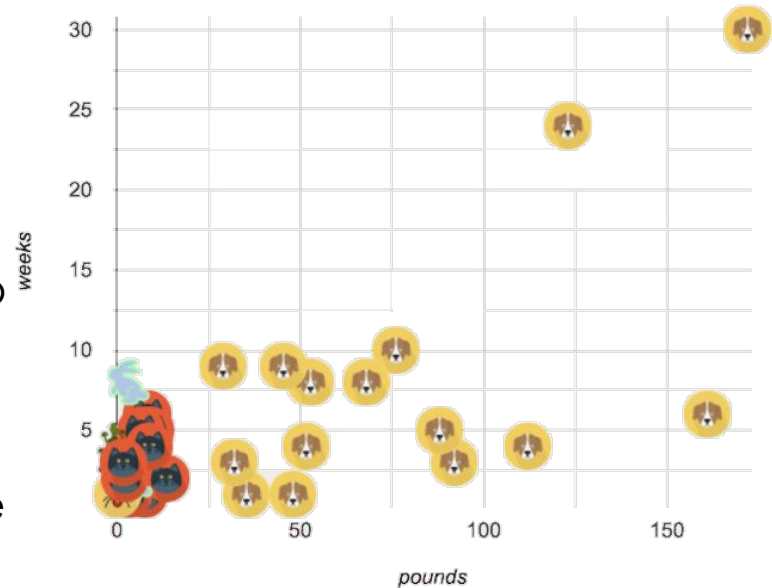




Problems with a Single Population

In groups of 3-4....

1. Does it make sense to analyze all the animals together? Why or why not?
2. When would it be important to break up the population into species-specific populations? What are they?
3. When would it be important to keep the whole population together?



Students, write your response!

Problems with a Single Population



Imagine that you've been handed a dataset from a country where half the people are rich and healthy, and the other half are poor and in poor health.

If we took a random sample of the population as a whole, we might think that they are 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

Ultimately, it might make more sense to ask 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.



Grouped Samples

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.



Grouped Samples

A “kitten” is an animal...
who *is a cat...* **and...** who *is young*.

How would you make a subset of just kittens?



Grouped Samples

1. Turn to [Grouped Samples from the Animals Dataset](#), and see what code will compute whether or not an animal is a kitten.
2. Can you fill in the code for the other subsets?
3. When you're done, open your saved Animals Starter File and type these definitions into the Definitions Area.





Grouped Samples

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)
```

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



Grouped Samples



How could we filter *and* sort a table?

How can we combine methods?



Displaying Samples

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!

Displaying Samples



Complete [Displaying Data](#), using what you've learned about subsets to make more sophisticated data displays.



Displaying Samples



Were any of your displays interesting or surprising?



Students, write your response!