

Name: _____



Student Workbook

Fall, 2022 - Pyret Edition



BOOTSTRAP
Equity • Scale • Rigor

Workbook v3.0

Brought to you by the Bootstrap team:

- Emmanuel Schanzer
- Kathi Fiser
- Shiriram Krishnamurthi
- Dorai Sitaram
- Joe Politz
- Ben Lerner
- Nancy Pfenning
- Flannery Denny
- Rachel Tabak

Visual Designer: Colleen Murphy

Bootstrap is licensed under a Creative Commons 3.0 Unported License. Based on a work from www.BootstrapWorld.org.
Permissions beyond the scope of this license may be available at contact@BootstrapWorld.org.

Computing Needs All Voices!

The pioneers pictured below are featured in our Computing Needs All Voices lesson. To learn more about them and their contributions, visit <https://bit.ly/bootstrap-pioneers>.



We are in the process of expanding our collection of pioneers. If there's someone else whose work inspires you, please let us know at <https://bit.ly/pioneer-suggestion>.

Notice and Wonder

Write down what you notice and wonder from the [What Most Schools Don't Teach](#) video.

"Notices" should be statements, not questions. What stood out to you? What do you remember? "Wonders" are questions.

What do you Notice?	What do you Wonder?

Reflection: Problem Solving Advantages of Diverse Teams

This reflection is designed to follow reading [LA Times Perspective: A solution to tech's lingering diversity problem? Try thinking about ketchup](#)

1) The author argues that tech companies with diverse teams have an advantage. Why?

2) What suggestions did the article offer for tech companies looking to diversify their teams?

3) What is one thing of interest to you in the author's bio?

4) Think of a time when you had an idea that felt out of the box. Did you share your idea? Why or why not?

5) Can you think of a time when someone else had a strategy or idea that you would never have thought of, but was interesting to you and/or pushed your thinking to a new level?

6) Based on your experience of exceptions to mainstream assumptions, propose another pair of questions that could be used in place of "Where do you keep your ketchup?" and "What would you reach for instead?".

Introduction to Computational Data Science

Many important questions (“What’s the best restaurant in town?”, “Is this law good for citizens?”, etc.) are answered with *data*. Data Scientists try and answer these questions by writing *programs that ask questions about data*.

Data of all types can be organized into **Tables**.

- Every Table has a **header row** and some number of **data rows**.
- **Quantitative data** is numeric and measures *an amount*, such as a person’s height, a score on a test, distance, etc. A list of quantitative data can be ordered from smallest to largest.
- **Categorical data** is data that specifies *qualities*, such as sex, eye color, country of origin, etc. Categorical data is not subject to the laws of arithmetic — for example, we cannot take the “average” of a list of colors.

The Animals Dataset

name	species	sex	age	fixed	legs	pounds	weeks
Sasha	cat	female	1	false	4	6.5	3
Snuffles	rabbit	female	3	true	4	3.5	8
Mittens	cat	female	2	true	4	7.4	1
Sunflower	cat	female	5	true	4	8.1	6
Felix	cat	male	16	true	4	9.2	5
Sheba	cat	female	7	true	4	8.4	6
Billie	snail	hermaphrodite	0.5	false	0	0.1	3
Snowcone	cat	female	2	true	4	6.5	5
Wade	cat	male	1	false	4	3.2	1
Hercules	cat	male	3	false	4	13.4	2
Toggle	dog	female	3	true	4	48	1
Boo-boo	dog	male	11	true	4	123	24
Fritz	dog	male	4	true	4	92	3
Midnight	dog	female	5	false	4	112	4
Rex	dog	male	1	false	4	28.9	9
Gir	dog	male	8	false	4	88	5
Max	dog	male	3	false	4	52.8	8
Nori	dog	female	3	true	4	35.3	1
Mr. Peanutbutter	dog	male	10	false	4	161	6
Lucky	dog	male	3	true	3	45.4	9
Kujo	dog	male	8	false	4	172	30
Buddy	lizard	male	2	false	4	0.3	3
Gila	lizard	female	3	true	4	1.2	4
Bo	dog	male	8	true	4	76.1	10
Nibblet	rabbit	male	6	false	4	4.3	2
Snuggles	tarantula	female	2	false	8	0.1	1
Daisy	dog	female	5	true	4	68	8
Ada	dog	female	2	true	4	32	3
Miaulis	cat	male	7	false	4	8.8	4
Heathcliff	cat	male	1	true	4	2.1	2
Tinkles	cat	female	1	true	4	1.7	3
Maple	dog	female	3	true	4	51.6	4

Categorical or Quantitative?

- **Quantitative data** measures an *amount* and can be ordered from smallest to largest.
- **Categorical data** specifies *qualities* and is not subject to the laws of arithmetic — for example, we cannot take the “average” of a list of colors.

Note: Numbers can be sometimes be categorical rather than quantitative!

For each piece of data below, circle whether it is **Categorical** or **Quantitative** data.

- | | | | |
|---|-------------|-------------|--------------|
| 1 | Hair color | categorical | quantitative |
| 2 | Age | categorical | quantitative |
| 3 | ZIP Code | categorical | quantitative |
| 4 | Year | categorical | quantitative |
| 5 | Height | categorical | quantitative |
| 6 | Sex | categorical | quantitative |
| 7 | Street Name | categorical | quantitative |

For each question, circle whether it will be answered by **Categorical** or **Quantitative** data.

- | | | | |
|----|---|-------------|--------------|
| 8 | We'd like to find out the average price of cars in a lot. | categorical | quantitative |
| 9 | We'd like to find out the most popular color for cars. | categorical | quantitative |
| 10 | We'd like to find out which puppy is the youngest. | categorical | quantitative |
| 11 | We'd like to find out which cats have been fixed. | categorical | quantitative |
| 12 | We want to know which people have a ZIP code of 02907. | categorical | quantitative |
| 13 | We'd like to sort a list of phone numbers by area code. | categorical | quantitative |

Questions and Column Descriptions

What questions can you ask about the animals dataset? For each question, can it be answered by this dataset? Make sure you have at least two questions that can be answered, and at least one that cannot.

What do you NOTICE about this dataset?	What do you WONDER about this dataset?	Answered by this dataset?
		Yes No

Describe the table, and two of the columns, by filling in the blanks below.

1. This dataset is _____, which contains _____ data rows.

2. Some of the columns are:

a. _____, which contains _____ data. Some example values are:

column name

categorical or quantitative

_____.

b. _____, which contains _____ data. Some example values are:

column name

categorical or quantitative

_____.

Introduction to Programming

The **Editor** is a software program we use to write Code. Our Editor allows us to experiment with Code on the right-hand side, in the **Interactions Area**. For Code that we want to *keep*, we can put it on the left-hand side in the **Definitions Area**. Clicking the "Run" button causes the computer to re-read everything in the Definitions Area and erase anything that was typed into the Interactions Area.

Data Types

Programming languages involve different data types, such as Numbers, Strings, Booleans, and even Images.

- Numbers are values like `1`, `0.4`, `1/3`, and `-8261.003`.
 - Numbers are *usually* used for quantitative data and other values are *usually* used as categorical data.
 - In Pyret, any decimal *must* start with a 0. For example, `0.22` is valid, but `.22` is not.
- Strings are values like `"Emma"`, `"Rosanna"`, `"Jen and Ed"`, or even `"08/28/1980"`.
 - All strings *must* be surrounded in quotation marks.
- Booleans are either `true` or `false`.

All values evaluate to themselves. The program `42` will evaluate to `42`, the String `"Hello"` will evaluate to `"Hello"`, and the Boolean `false` will evaluate to `false`.

Operators

Operators (like `+`, `-`, `*`, `<`, etc.) work the same way in Pyret that they do in math.

- Operators are written between values, for example: `4 + 2`.
- In Pyret, operators must always have a space around them. `4 + 2` is valid, but `4+2` is not.
- If an expression has different operators, parentheses must be used to show order of operations. `4 + 2 + 6` and `4 + (2 * 6)` are valid, but `4 + 2 * 6` is not.

Applying Functions

Applying functions works much the way it does in math. Every function has a name, takes some inputs, and produces some output. The function name is written first, followed by a list of arguments in parentheses.

- In math this could look like $f(5)$ or $g(10, 4)$.
- In Pyret, these examples would be written as `f(5)` and `g(10, 4)`.
- Applying a function to make images would look like `star(50, "solid", "red")`.
- There are many other functions, for example `num-sqr`, `num-sqrt`, `triangle`, `square`, `string-repeat`, etc.

Functions have contracts, which help explain how a function should be used. Every contract has three parts:

- The *Name* of the function - literally, what it's called.
- The *Domain* of the function - what *types of values* the function consumes, and in what order.
- The *Range* of the function - what *type of value* the function produces.

Numbers and Strings

Make sure you've loaded the [code.pyret.org\(CPO\)](http://code.pyret.org(CPO)), clicked "Run", and are working in the Interactions Area.

Numbers

1) Try typing `42` into the Interactions Area and hitting "Enter". What is the largest number the editor can handle?

2) Try typing `0.5`. Then try typing `.5`. Then try clicking on the answer. Experiment with other decimals. Explain what you understand about how decimals work in this programming language. _____

3) What happens if you try a fraction like `1/3`? _____

4) Try writing **negative** integers, fractions and decimals. What do you learn? _____

Strings

String values are always in quotes.

5) Is `42` the same as `"42"`? Why or why not? Write your answer below:

- Try typing your name (*in quotes!*).
- Try typing a sentence like "I'm excited to learn to code!" (*in quotes!*).
- Try typing your name with the opening quote, but *without the closing quote*. Read the error message!
- Now try typing your name *without any quotes*. Read the error message!

6) Explain what you understand about how strings work in this programming language. _____

Operators

7) Just like math, Pyret has operators like `+`, `-`, `*` and `/`. Try typing in `4 + 2`, and then `4+2` (without the spaces). What can you conclude from this?

8) Type in the following expressions, **one at a time**: `4 + 2 * 6`, `(4 + 2) * 6`, `4 + (2 * 6)`. What do you notice? _____

9) Try typing in `4 + "cat"`, and then `"dog" + "cat"`. What can you conclude from this?

Booleans

Boolean-producing expressions are yes-or-no questions and will always evaluate to either **true** ("yes") or **false** ("no"). What will each of the expressions below evaluate to? Write down your prediction in the blanks provided and then type the code into the Interactions Area to see what it returns.

	Prediction	Result		Prediction	Result
1) <code>3 <= 4</code>	_____	_____	2) <code>"a" > "b"</code>	_____	_____
3) <code>3 == 2</code>	_____	_____	4) <code>"a" < "b"</code>	_____	_____
5) <code>2 < 4</code>	_____	_____	6) <code>"a" == "b"</code>	_____	_____
7) <code>5 >= 5</code>	_____	_____	8) <code>"a" <> "a"</code>	_____	_____
9) <code>4 >= 6</code>	_____	_____	10) <code>"a" >= "a"</code>	_____	_____
11) <code>3 <> 3</code>	_____	_____	12) <code>"a" <> "b"</code>	_____	_____

13) In your own words, describe what `<` does.

14) In your own words, describe what `>=` does.

15) In your own words, describe what `<>` does.

	Prediction:	Result:
16) <code>string-contains("catnap", "cat")</code>	_____	_____
17) <code>string-contains("cat", "catnap")</code>	_____	_____

18) How many **Numbers** are there in the entire universe?

19) How many **Strings** are there in the entire universe?

20) How many **Booleans** are there in the entire universe?

Applying Functions

Type this line of code into the Interactions Area and hit "Enter":

```
triangle(50, "solid", "red")
```

- 1) What is the name of this function? _____
- 2) What did the expression evaluate to? _____
- 3) How many arguments does `triangle` expect? _____
- 4) What data type does the `triangle` function produce? _____

Catching Bugs

The following lines of code are all BUGGY! Read the code and the error messages to identify the mistake.

5) `triangle(20, "solid" "red")`

```
Pyret didn't understand your program around
triangle(20, "solid" "red")
```

Can you spot the mistake? _____

6) `triangle(20, "solid")`

This application expression errored:

```
triangle(20, "solid")
```

2 arguments were passed to the **operator**. The **operator** evaluated to a function accepting 3 parameters. An application expression expects the number of parameters and arguments to be the same.

Can you spot the mistake? _____

7) `triangle(20, 10, "solid", "red")`

This application expression errored:

```
triangle(20, 10, "solid", "red")`
```

4 arguments were passed to the **operator**. The **operator** evaluated to a function accepting 3 parameters. An application expression expects the number of parameters and arguments to be the same.

Can you spot the mistake? _____

8) `triangle (20, "solid", "red")`

Pyret thinks this code is probably a function call:

```
triangle (20, "solid", "red")
```

Function calls must not have space between the **function expression** and the arguments.

Can you spot the mistake? _____

Frayer Model: Domain and Range

My Definition	Facts and Characteristics
Domain	
Examples	Non-Examples
My Definition	Facts and Characteristics
Range	
Examples	Non-Examples

Practicing Contracts: Domain & Range

Consider the following contract:

```
is-beach-weather :: Number, String -> Boolean
```

- 1) What is the **Name** of this function? _____
- 2) How many arguments are in this function's **Domain**? _____
- 3) What is the **Type** of this function's **first argument**? _____
- 4) What is the **Type** of this function's **second argument**? _____
- 5) What is the **Range** of this function? _____
- 6) Circle the expression below that shows the correct application of this function, based on its contract.
 - A. `is-beach-weather(70, 90)`
 - B. `is-beach-weather(80, 100, "cloudy")`
 - C. `is-beach-weather("sunny", 90)`
 - D. `is-beach-weather(90, "stormy weather")`

Consider the following contract:

```
cylinder :: Number, Number, String -> Image
```

- 7) What is the **Name** of this function? _____
- 8) How many arguments are in this function's **Domain**? _____
- 9) What is the **Type** of this function's **first argument**? _____
- 10) What is the **Type** of this function's **second argument**? _____
- 11) What is the **Type** of this function's **third argument**? _____
- 12) What is the **Range** of this function? _____
- 13) Circle the expression below that shows the correct application of this function, based on its contract.
 - A. `cylinder("red", 10, 60)`
 - B. `cylinder(30, "green")`
 - C. `cylinder(10, 25, "blue")`
 - D. `cylinder(14, "orange", 25)`

Matching Expressions and Contracts

Match the contract (left) with the expression described by the function being used (right).

Contract		Expression
<code># make-id :: String, Number -> Image</code>	1	A <code>make-id("Savannah", "Lopez", 32)</code>
<code># make-id :: String, Number, String -> Image</code>	2	B <code>make-id("Pilar", 17)</code>
<code># make-id :: String -> Image</code>	3	C <code>make-id("Akemi", 39, "red")</code>
<code># make-id :: String, String -> Image</code>	4	D <code>make-id("Raïssa", "McCracken")</code>
<code># make-id :: String, String, Number -> Image</code>	5	E <code>make-id("von Einsiedel")</code>

Contract		Expression
<code># is-capital :: String, String -> Boolean</code>	6	A <code>show-pop("Juneau", "AK", 31848)</code>
<code># is-capital :: String, String, String -> Boolean</code>	7	B <code>show-pop("San Juan", 395426)</code>
<code># show-pop :: String, Number -> Image</code>	8	C <code>is-capital("Accra", "Ghana")</code>
<code># show-pop :: String, String, Number -> Image</code>	9	D <code>show-pop(3751351, "Oklahoma")</code>
<code># show-pop :: Number, String -> Number</code>	10	E <code>is-capital("Albany", "NY", "USA")</code>

Using Contracts

`ellipse :: Number, Number, String, String -> Image`



Use the contract to write an expression that generates a similar image:



Use the contract to write an expression that generates a similar image:

What changes with the first Number?

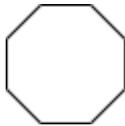
What about the shape changes with the second Number?

Write an expression using `ellipse` to produce a circle.

`regular-polygon :: Number, Number, String, String -> Image`



Use the contract to write an expression that generates a similar image:



Use the contract to write an expression that generates a similar image:

What changes with the first Number?

What about the shape changes with the second Number?

Use `regular-polygon` to write an expression for a square!

How would you describe a **regular polygon** to a friend?

Triangle Contracts

1) What kind of triangle does the `triangle` function produce? _____

There are lots of other kinds of triangles! And Pyret has lots of other functions that make triangles!

```
triangle :: (size:: Number, style :: String, color :: String) -> Image
```

```
right-triangle :: (base::Number, height::Number, style::String, color::String) -> Image
```

```
isosceles-triangle :: (leg::Number, angle::Number, style::String, color::String) -> Image
```

2) Why do you think `triangle` only needs one number, while `right-triangle` and `isosceles-triangle` need two numbers and `triangle-sas` needs three?

3) Write `right-triangle` expressions for the images below. *One argument for each should be 100.*





4) What do you think the numbers in `right-triangle` represent? _____

5) Write `isosceles-triangle` expressions for the images below. *1 argument for each should be 100.*





6) What do you think the numbers in `isosceles-triangle` represent?

7) Write 2 expressions that would build **right-isosceles** triangles. Use `right-triangle` for one expression and `isosceles-triangle` for the other expression.



Radial Star

```
radial-star :: (  
  points :: Number,  
  inner-radius :: Number,  
  full-radius :: Number,  
  style :: String,  
  color :: String  
) -> Image
```

Using the detailed contract above, match each image to the expression that describes it.

Image			Expression
	1	A	<code>radial-star(5, 50, 200, "solid", "black")</code>
	2	B	<code>radial-star(7, 100, 200, "solid", "black")</code>
	3	C	<code>radial-star(7, 100, 200, "outline", "black")</code>
	4	D	<code>radial-star(10, 150, 200, "solid", "black")</code>
	5	E	<code>radial-star(10, 20, 200, "solid", "black")</code>
	6	F	<code>radial-star(100, 20, 200, "outline", "black")</code>
	7	G	<code>radial-star(100, 100, 200, "outline", "black")</code>

Displaying Categorical Data

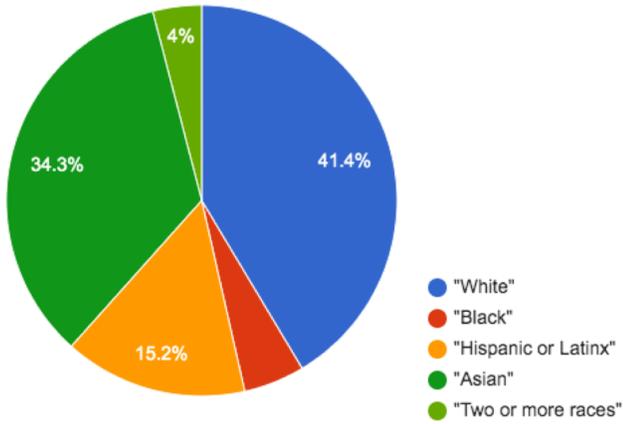
Data Scientists use **displays** to visualize data. You've probably seen some of these charts, graphs and plots yourselves! When it comes to displaying **Categorical Data**, there are two displays that are especially useful.

1. **Bar charts** show the *count or percentage* of rows in each category.
 - Bar charts provide a visual representation of the frequency of values in a categorical column.
 - Bar charts have a bar for every category in a column.
 - The more rows in a category, the taller the bar.
 - Bars in a bar chart can be shown in *any order*, without changing the meaning of the chart. However, bars are usually shown in some sensible order (bars for the number of orders for different t-shirt sizes might be presented in order of smallest to largest shirt).
2. **Pie charts** show the *percentage* of rows in each category.
 - Pie charts provide a visual representation of the relative frequency of values in a categorical column.
 - Pie charts have a slice for every category in a column.
 - The more rows in a category, the larger the slice.
 - Slices in a pie chart can be shown in *any order*, without changing the meaning of the chart. However, slices are usually shown in some sensible order (e.g. slices might be shown in alphabetical order or from the smallest to largest slice).

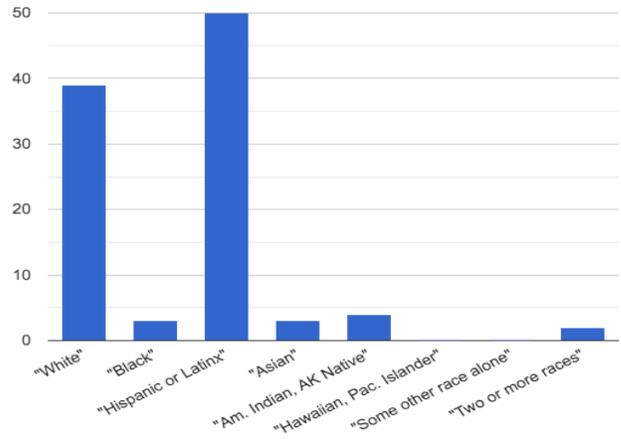
Bar & Pie Chart - Notice and Wonder

What do you notice and wonder about the displays below?

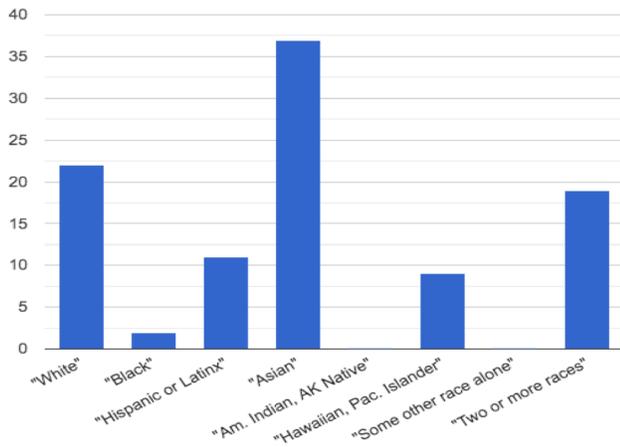
San Francisco Unified School District, CA



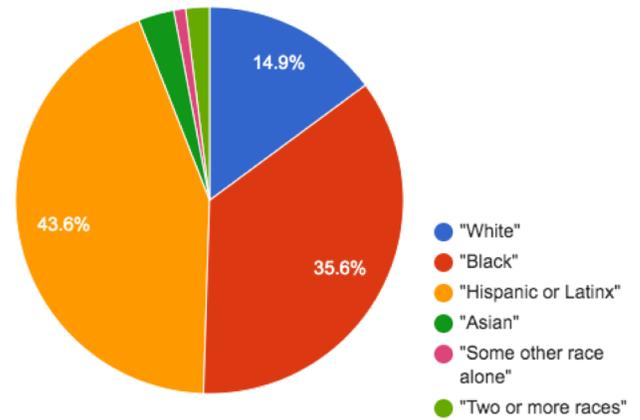
Albuquerque Public Schools, NM



Hawaii DOE, HI



Hartford School District, CT



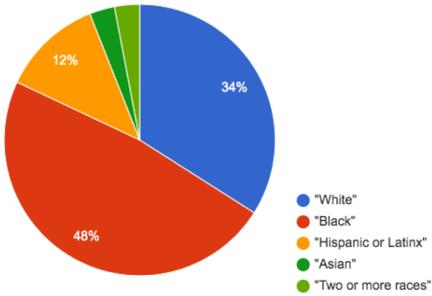
What do you Notice?

What do you Wonder?

Matching Bar Charts to Pie Charts

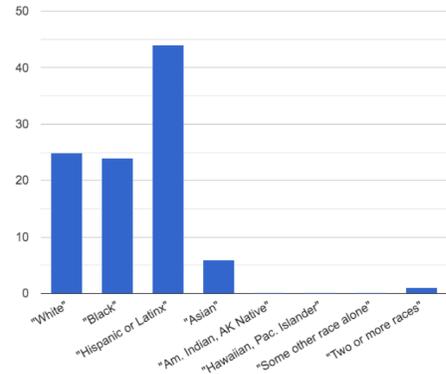
Match each bar chart below to the pie chart that displays the racial demographic data from the same school district.

Cleveland Municipal School District

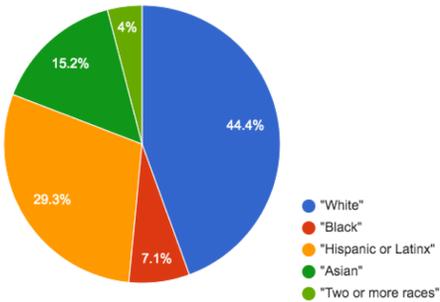


1

A

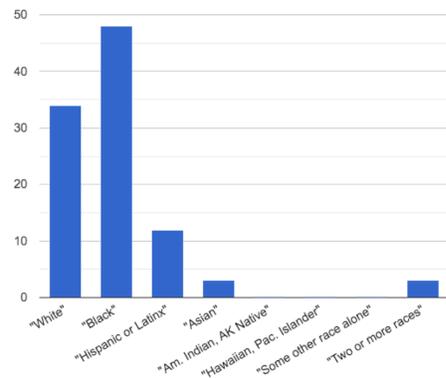


San Diego City Unified School District

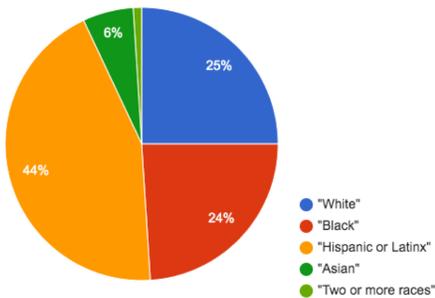


2

B

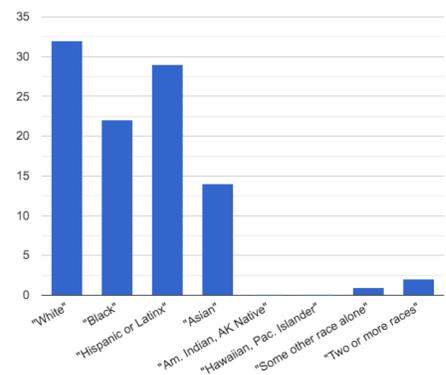


Houston Independent School District

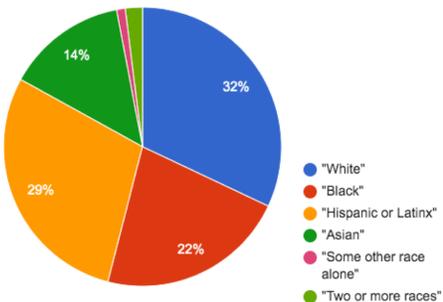


3

C

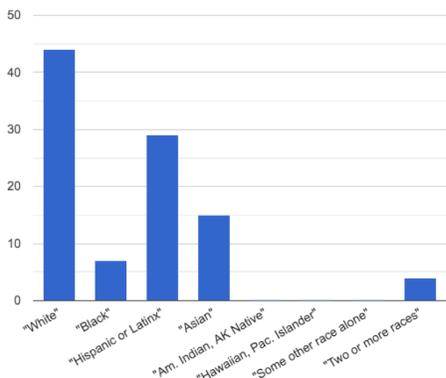


New York City Dept of Education



4

D



Exploring Displays

Using your Contracts page and the Animals Starter File, make each type of display below in Pyret. Then sketch the displays and answer the questions. Be sure to add examples of the code you use to your contracts page!

Pie Charts

Sketch a pie chart below

Displays 1 column(s) of category data.
What does this display tell us?

Bar Charts

Sketch a bar chart below

Displays _____ column(s) of _____ data.
What does this display tell us?

Box Plots

Sketch a box plot below

Displays _____ column(s) of _____ data.
What do you think this display tells us?

Histograms

Sketch a histogram below

Displays _____ column(s) of _____ data.
What do you think this display tells us?

(More) Exploring Displays

For each type of display, fill in the information below.

Scatter Plot	Linear Regression Plot
<p data-bbox="315 268 591 296">Sketch a scatter plot below</p> <p data-bbox="110 751 748 808">Displays _____ column(s) of _____ data. What do you think this display tells us?</p> <hr data-bbox="110 871 790 875"/> <hr data-bbox="110 940 790 945"/> <hr data-bbox="110 1010 790 1014"/> <hr data-bbox="110 1079 790 1083"/>	<p data-bbox="980 268 1354 296">Sketch a linear regression plot below</p> <p data-bbox="824 751 1463 808">Displays _____ column(s) of _____ data. What do you think this display tells us?</p> <hr data-bbox="824 871 1505 875"/> <hr data-bbox="824 940 1505 945"/> <hr data-bbox="824 1010 1505 1014"/> <hr data-bbox="824 1079 1505 1083"/>

What Display Goes with Which Kind of Data?

Match the Display with the description of the data being plotted. Some descriptions may go with more than one display!

Pie Charts 1

A 1 column of Quantitative Data

Bar Charts 2

Histograms 3

B 2 columns of Quantitative Data

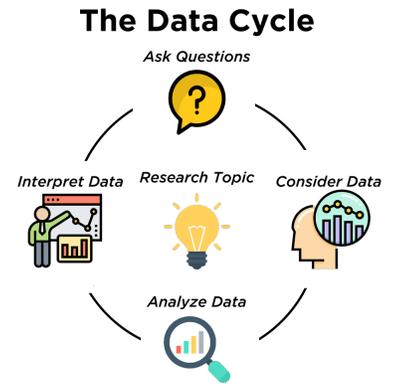
Box Plots 4

Scatter Plots 5

C 1 column of Categorical Data

The Data Cycle

Data Science is all about *asking questions of data*. Sometimes the answer is easy to compute. Sometimes the answer to a question is *already in the dataset* - no computation needed. And sometimes the answer just sparks more questions! Data Scientists ask a lot of questions, and each one adds a chapter to the story of their research. Even if a question is a "dead-end", it's valuable to share what the question was and what work you did to answer it!



We start by **Asking Questions** - questions that can be answered with data. It's useful to think of all questions as falling into four categories:

- **Lookup questions** can be answered simply by looking up a single value in the table. Examples of lookup questions might be "How many legs does Felix have?" or "What species is Sheba?"
- **Arithmetic questions** can be answered by computing an answer across a single column. Examples of arithmetic questions might be "How much does the heaviest animal weigh?" or "What is the average age of animals from the shelter?"
- **Statistical questions** are where things get interesting! The answers are never black-and-white. If we asked, "How old are animals at the shelter?" we could report back the average age, the age that shows up most frequently or the range of the ages. There are also some statistical questions that deal with *relationships between two columns*: "Do cats tend to be adopted faster than dogs?" or "Are older animals heavier than young ones?"
- **Questions we can't answer** are pretty common, too! We might wonder where the animal shelter is located, or what time of year the data was gathered! But the data in the table won't help us answer that question, so as Data Scientists we might need to do some research *beyond the data*. And if nothing turns up, we simply recognize that there are limits to what we can analyze.

Next, we **Consider Data**. This could be done by conducting a survey, observing and recording data, or finding a dataset that meets our needs. Since our data is contained in a table, it's useful to start by asking two questions:

- What rows do we care about? - Is it all the animals? Just the lizards?
- What columns do we need? - Are we examining the ages of the animals? Their weights?

Then, we **Analyze the Data**, by producing data displays and new tables of filtered or transformed data and using them to identify patterns and relationships.

- Are we making a pie chart? A bar chart? Something else?

Finally, we **Interpret the Data**, by answering our questions and summarizing the results. As we've already seen from the Animals Dataset, these interpretations often lead to *new questions*... and the cycle begins again.

Which Question Type?

Take a moment to look at the table of Pokemon data below:

name	type1	hitpoint	attack	defense	speed
Bulbasaur	Grass	45	49	49	45
Ivysaur	Grass	60	62	63	60
Venusaur	Grass	80	82	83	80
Mega Venusaur	Grass	80	100	123	80
Charmander	Fire	39	52	43	65
Charmeleon	Fire	58	64	58	80
Charizard	Fire	78	84	78	100
Mega Charizard X	Fire	78	130	111	100
Mega Charizard Y	Fire	78	104	78	100
Squirtle	Water	44	48	65	43
Wartortle	Water	59	63	80	58

For each question about this data below, write down whether it's a Lookup, Arithmetic, or Statistical Question.

	Question	Type	Which Rows?	Column(s)?
1	What type is Charizard?			
2	Which Pokemon is the fastest?			
3	What is Wartortle's attack score?			
4	What is the average defense score?			
5	What is a typical defense score?			
6	Is Ivy faster than Venusaur?			
7	Does speed affect attack score?			
8	What is the most common type?			
9	Does one type tend to be faster than others?			
10	Are hit-points (hp) similar for all Pokemon?			
11	Which Pokemon has a speed of 43?			

Data Cycle: Consider Data

For each question about the animals dataset below, identify the type of question and fill in the Rows and Columns needed to answer it.

<p>Ask Questions</p> 	<p><i>How old is Boo-boo?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>

<p>Ask Questions</p> 	<p><i>Are there more cats than dogs in the shelter?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>

<p>Ask Questions</p> 	<p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>

<p>Ask Questions</p> 	<p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>

Data Cycle: Analyzing with Displays

For each question below, complete the first three steps of the Data Cycle. Once you know what code to write, type it into Pyret and try it out!

Ask Questions 	<i>How many of each species are at the shelter?</i> What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	What code will make the table or display you want? _____ _____

Ask Questions 	<i>How many of each sex are at the shelter?</i> What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	What code will make the table or display you want? _____ _____

Ask Questions 	What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	What code will make the table or display you want? _____ _____

Probability, Inference, and Sample Size

How can you tell if a coin is fair, or designed to cheat you? Statisticians know that a fair coin should turn up "heads" about as often as "tails", so they begin with the **null hypothesis**: they assume the coin is fair, and start flipping it over and over to record the results.

A coin that comes up "heads" three times in a row could still be fair! The odds are 1-in-8, so it's totally possible that the null hypothesis is still true. But what if it comes up "heads" five times in a row? Ten times?

Eventually, the chances of the coin being fair get smaller and smaller, and a Data Scientist can say "this coin is a cheat! The chances of it being fair are one in a million!"

By sampling the flips of a coin, we can *infer* whether the coin itself is fair or not. **Inference** plays a major role in Data Science and Statistics!

- If we survey pet owners about whether they prefer cats or dogs, the **null hypothesis** is that the odds of someone preferring dogs are about the same as them preferring cats. And if the first three people we ask vote for dogs (a 1-in-8 chance), the null hypothesis could still be true! But after five people? Ten?
- If we're looking for gender bias in hiring, we might start with the null hypothesis that no such bias exists. If the first three people hired are all men, that doesn't necessarily mean there's a bias! But if 30 out of 35 hires are male, this is evidence that undermines the null hypothesis and suggests a real problem.
- If we poll voters for the next election, the **null hypothesis** is that the odds of voting for one candidate are the same as voting for the other. But if 80 out of 100 people say they'll vote for the same candidate, we might reject the null hypothesis and infer that the population as a whole is biased towards that candidate!

Sample size matters! The more bias there is, the smaller the sample we need to detect it. Major biases might need only a small sample, but subtle ones might need a huge sample to be found. However, choosing a **good sample** can be tricky!

Random Samples are a subset of a population in which each member of the subset has an equal chance of being chosen. A random sample is intended to be a representative subset of the population. The larger the random sample, the more closely it will represent the population and the better our inferences about the population will tend to be.

Grouped Samples are a subset of a population in which each member of the subset was chosen for a specific reason. For example, we might want to look at the difference in trends between two groups ("Is the age of a dog a bigger factor in adoption time v. the age of a cat?"). This would require making grouped samples of *just the dogs* and *just the cats*.

Finding the Trick Coin

There are three coins defined in the [Fair Coins Starter File](#): `coin1` , `coin2` , and `coin3` .

One of these coins is fair, one will land on "heads" 75% of the time, and one will land on "heads" 90% of the time. **Which one is which?**

1) In the table below, record five flips of each coin and the number of "heads" you saw. Then convert that number into a *percentage* . Finally, write a "Y" or "N" depending on whether you think each coin is fair based on your sample.

Sample	coin1	coin2	coin3
1			
2			
3			
4			
5			
#heads	/5	/5	/5
% heads	%	%	%
fair?			

2) In the table below, record 15 more flips of each coin and the number of "heads" you saw. Then convert that number into a *percentage* . Finally, write a "Y" or "N" depending on whether you think each coin is fair based on your sample.

Sample	coin1	coin2	coin3
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
#heads	/20	/20	/20
% heads	%	%	%
fair?			

3) Which weighted coin did you identify first: fair, 75%, or 90%? _____

4) Why was that coin the easiest to find? _____

Sampling and Inference

You'll need to open the [Expanded Animals Starter File](#), and save a copy to complete this worksheet.

1) Evaluate the `more-animals` table in the Interactions Area. This is the *complete* population of animals from the shelter! Below is a true statement about that population:

The population is 47.7% fixed and 52.3% unfixed.

Type each of the following lines into the Interactions Area and hit "Enter".

```
random-rows(more-animals, 10)
```

```
random-rows(more-animals, 40)
```

2) What do you get? _____

3) What is the contract for `random-rows`? _____

4) What does the `random-rows` function do? _____

5) In the Definitions Area, define `small-sample` and `large-sample` to be `random-rows(more-animals, 10)` and `random-rows(more-animals, 40)`, respectively.

6) Make a `pie-chart` for the animals in each sample, showing percentages of fixed and unfixed.

- The percentage of fixed animals in the entire population is 47.7%
- The percentage of fixed animals in `small-sample` is _____
- The percentage of fixed animals in `large-sample` is _____

7) Make a `pie-chart` for the animals in each sample, showing percentages for each species.

- The percentage of tarantulas in the entire population is roughly 5%
- The percentage of tarantulas in `small-sample` is _____
- The percentage of tarantulas in `large-sample` is _____

8) Click "Run" to direct the computer to generate a different set of random samples of these sizes. Make a new `pie-chart` for each sample, showing percentages for each species.

- The percentage of tarantulas in the entire population is roughly 5%
- The percentage of tarantulas in `small-sample` is _____
- The percentage of tarantulas in `large-sample` is _____

9) Which sample size gave us a more accurate inference about the whole population? Why?

Choosing Your Dataset

When selecting a dataset to explore, *pick something that matters to you!* You'll be working with this data for a while, so you don't want to pick something at random just to get it done.

When choosing a dataset, it's a good idea to consider a few factors:

1. Is it **interesting**? Pick a dataset you're genuinely interested in, so that you can explore questions that matter to you!
2. Is it **relevant**? Pick a dataset that deals with something personally relevant to you and your community! Does this data impact you in any way? Are there questions you have about the dataset that mean something to you or someone you know?
3. Is it **familiar**? Pick a dataset you know about, so you can use your expertise to deepen your analysis! You wouldn't be able to make samples of the Animals Dataset properly if you didn't know that some animals are much bigger or longer-lived than others.

Consider and Analyze

Fill in the tables below by considering the rows and columns you need. Look up the Contract for the display and record the Pyret code you'd need to make it. If time allows, type your code into the editor to see your display! *The first column has been filled in for you.*

1) A pie-chart showing the `species` of animals from the shelter.

Which Rows?	Which Column(s)?	What will you Create?
<i>All the animals</i>		

code: _____

2) A bar-chart showing the `sex` of animals from the shelter.

Which Rows?	Which Column(s)?	What will you Create?
<i>All the animals</i>		

code: _____

3) A histogram of the number of `pounds` that animals weigh.

Which Rows?	Which Column(s)?	What will you Create?
<i>All the animals</i>		

code: _____

4) A box-plot of the number of `pounds` that animals weigh.

Which Rows?	Which Column(s)?	What will you Create?
<i>All the animals</i>		

code: _____

5) A scatter-plot, using the animals' `species` as the labels, `age` as the x-axis, and `pounds` as the y-axis.

Which Rows?	Which Column(s)?	What will you Create?
<i>All the animals</i>		

code: _____

6) A scatter-plot, using the animals' `name` as the labels, `pounds` as the x-axis, and `weeks` as the y-axis.

Which Rows?	Which Column(s)?	What will you Create?
<i>All the animals</i>		

code: _____

My Dataset

The _____ dataset contains _____ data rows.

1) I'm interested in this data because _____

2) My friends, family or neighbors would be interested because _____

3) Someone else should care about this data because _____

4) In the table below, write down what you Notice and Wonder about this dataset.

What do you NOTICE?	What do you WONDER?	Question
		Lookup Arithmetic Statistical Can't Answer

5) Look at each Wonder you wrote above. Is it an arithmetic question? A question you can't answer from the data? Circle the type of question. Choose two columns to describe below

6) _____, which contains _____ data. Example values from this column include:
column name categorical/quantitative

7) _____, which contains _____ data. Example values from this column include:
column name categorical/quantitative

Data Cycle: Shape of My Dataset

Use the Data Cycle to explore the distribution of one or more categorical columns using **pie-charts** and **bar-charts**, and write down your findings.

Ask Questions 	What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	What code will make the table or display you want? _____ _____
Interpret Data 	What did you find out? What can you infer? _____ _____ What new question(s) does this raise? _____ _____ _____

Ask Questions 	What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	What code will make the table or display you want? _____ _____
Interpret Data 	What did you find out? What can you infer? _____ _____ What new question(s) does this raise? _____ _____ _____

Histograms

To best understand histograms, it's helpful to contrast them first with bar charts.

Bar charts show the number of rows belonging to a given category. The more rows in each category, the taller the bar.

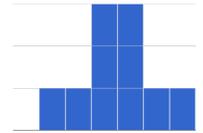
- *Bar charts provide a visual representation of the frequency of values in a **categorical** column.*
- There's no strict numerical way to order these bars, but **sometimes there's an order** that makes sense. For example, bars for the sales of different t-shirt sizes might be presented in order of smallest to largest shirt.

Histograms show the number of rows that fall within certain intervals, or "bins", on a horizontal axis. The more rows that fall within a particular "bin", the taller the bar.

- *Histograms provide a visual representation of the frequencies (or relative frequencies) of values in a **quantitative** column.*
- Quantitative data **can always be ordered**, so the bars of a histogram always progress from smallest (on the left) to largest (on the right).
- When dealing with histograms, it's important to select a good **bin size**. If the bins are too small or too large, it is difficult to see the shape of the dataset. Choosing a good bin size can take some trial and error!

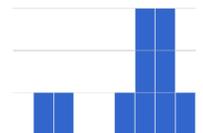
The **shape** of a dataset tells us which values are more or less common.

- In a **symmetric** dataset, values are just as likely to occur a certain distance above the mean as below the mean. Each side of a symmetric distribution looks almost like a mirror-image of the other.

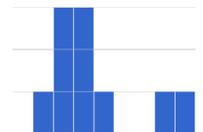


- Some extreme values may be far greater or far lower than the other values in a dataset. These extreme values are called **outliers**.

- A dataset that is **skewed left** has a few values that are unusually low. The histogram for a skewed left dataset has a few data points that are stretched out to the left (lower) end of the x-axis.



- A dataset that is **skewed right** has a few values that are unusually high. The histogram for a skewed right dataset has a few data points that are stretched out to the right (higher) end of the x-axis.



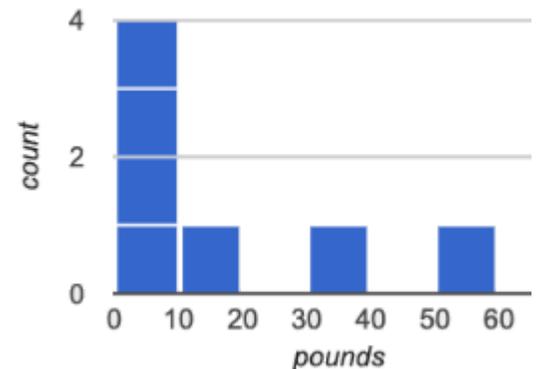
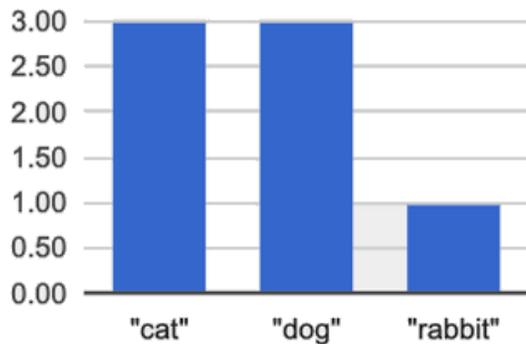
- One way to visualize the difference between a histogram of data that is **skewed left** or **skewed right** is to think about the lengths of our toes on our left and right feet. Much like the bar lengths of a histogram that is "skewed left", our left feet have smaller toes on the left and a bigger toe on the right. Our right feet have the big toe on the left and smaller toes on the right, more closely resembling the shape of a histogram of "skewed right" data.

Summarizing Columns

name	species	age	pounds
"Sasha"	"cat"	1	6.5
"Boo-boo"	"dog"	11	12.3
"Felix"	"cat"	16	9.2
"Nori"	"dog"	6	35.3
"Wade"	"cat"	1	3.2
"Nibblet"	"rabbit"	6	4.3
"Maple"	"dog"	3	51.6

1	How many cats are there in the table above?	
2	How many dogs are there?	
3	How many animals weigh between 0-20 pounds?	
4	How many animals weigh between 20-40 pounds?	
5	Are there more animals weighing 40-60 than 60-140 pounds?	

The two displays below both summarize this table. The display on the left is a **Bar Chart**, the one on the right is a **Histogram**. What is similar about them? What is different?



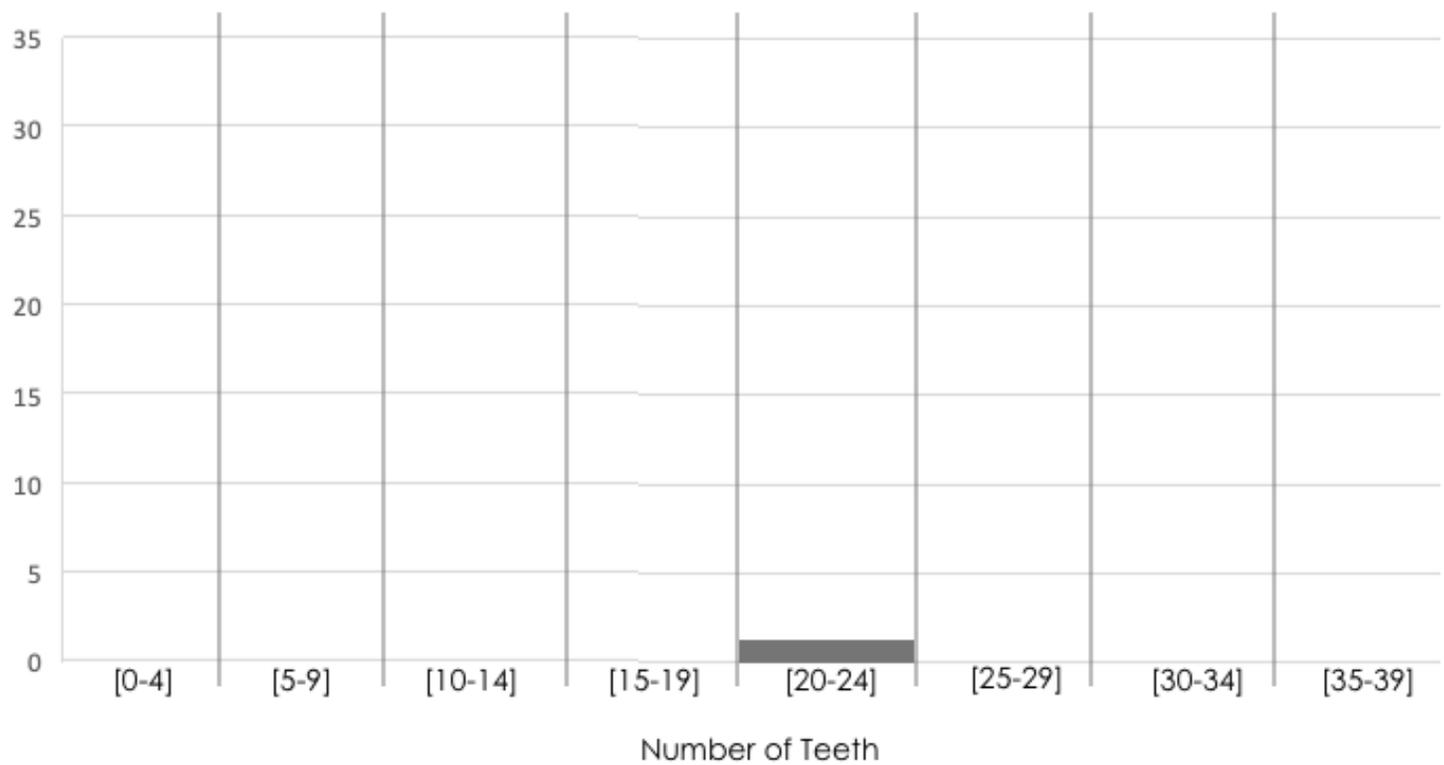
Similarities	Differences

Making Histograms

Suppose we have a dataset for a group of 50 adults, showing the number of teeth each person has:

Number of teeth	Count
0	5
22	1
26	1
27	1
28	4
29	3
30	5
31	3
32	27

Draw a histogram for the table in the space below. For each row, find which interval (or "bin") on the x-axis represents the right number of teeth. Then fill in the box so that the height of the box is equal to the *sum of the counts* that fit into that interval. One of the intervals has been completed for you.



Reading Histograms

Students watched 5 videos, and rated them on a scale of 1 to 10. While the **average score** for every video is the same (5.5), the **shapes** of the ratings distributions were very different! *Match* the summary description (left) with the *shape* of the histogram of student ratings (right). For each histogram, **the x-axis is the score, and the y-axis is the number of students who gave it that score**. These axes are intentionally unlabeled - focusing on the *shape* is what matters here!

Most of the students were fine with the video, but a couple of them gave it an unusually low rating.

1

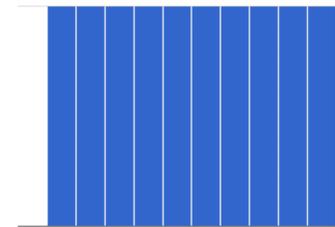
A



Most of the students were okay with the video, but a couple students gave it an unusually high rating.

2

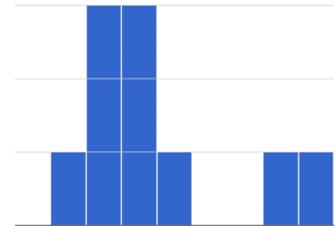
B



Students tended to give the video an average rating, and they weren't likely to stray far from the average.

3

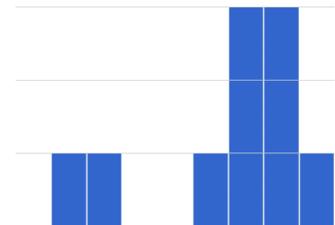
C



Students either really liked or really disliked the video.

4

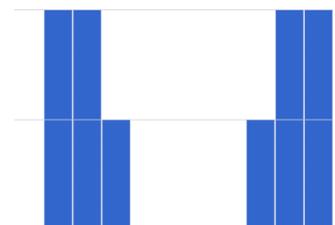
D



Reactions to the video were all over the place: high ratings and low ratings and inbetween ratings were all equally likely.

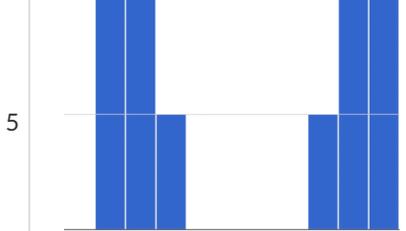
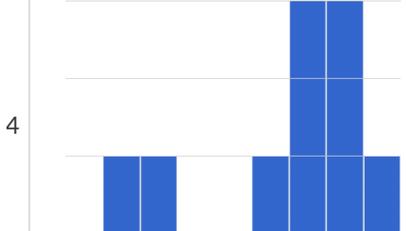
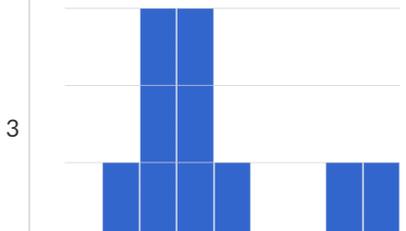
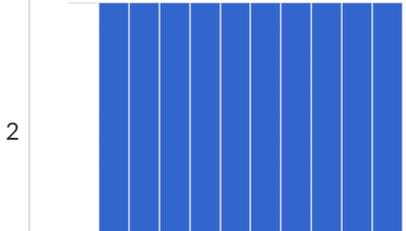
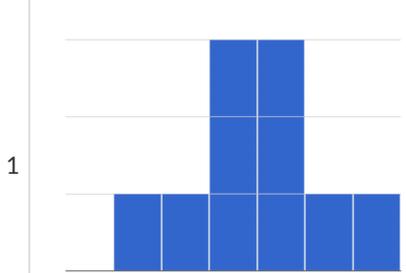
5

E



Identifying Shape - Histograms

Describe the shape of the histograms on the left. Do your best to incorporate the vocabulary you've been introduced to.



Data Cycle: Shape of the Animals Dataset

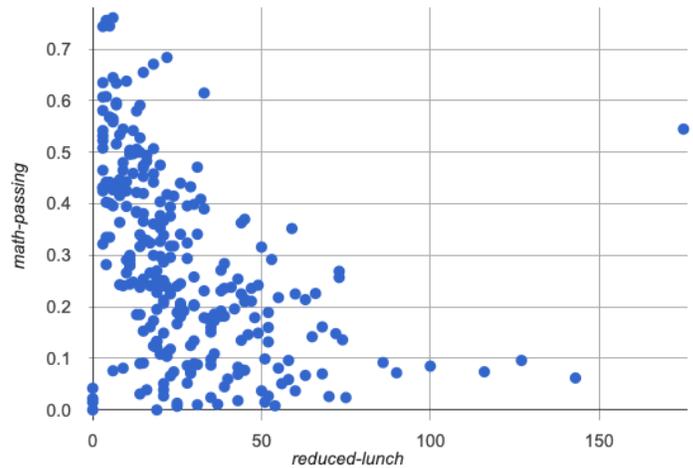
Describe two **histograms** made from columns of the animals dataset. The first question is given - you'll need to come up with the second question on your own!

<p>Ask Questions</p> 	<p><i>What is the distribution of weight among all animals at the shelter?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>
<p>Analyze Data</p> 	<p>What code will make the table or display you want?</p> <hr/>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/>

<p>Ask Questions</p> 	<p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>
<p>Analyze Data</p> 	<p>What code will make the table or display you want?</p> <hr/>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/>

Outliers: Should they Stay or Should they Go?

Tahli and Fernando are looking at a scatter plot showing the relationship between poverty and test scores at schools in Michigan. They find a trend, with low-poverty schools generally having higher test scores than high-poverty schools. However, one school is an extreme outlier: the highest poverty school in the state also has higher test scores than most of the other schools!



Tahli thinks the outlier should be removed before they start analyzing, and Fernando thinks it should stay. Here are their reasons:

Tahli's Reasons:	Fernando's Reasons:
<p>This outlier is so far from every other school - it <i>has</i> to be a mistake. Maybe someone entered the poverty level or the test scores incorrectly! We don't want those errors to influence our analysis. Or maybe it's a magnet, exam or private school that gets all the top-performing students. It's not right to compare that to non-magnet schools.</p>	<p>Even if it's a magnet, those are still Michigan students and their data should not be erased. Maybe it's not a mistake or a special school! Maybe the school has an amazing new strategy that's different from other schools! Instead of removing an inconvenient data point from the analysis, we should be focusing our analysis on what is happening there.</p>

1) Do you think this outlier should stay or go? Why?

2) What information would help you make your decision?

Data Cycle: Shape of My Dataset

Use the Data Cycle to explore the distribution of one or more quantitative columns using **histograms**, and write down your findings.

Ask Questions 	What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	What code will make the table or display you want? _____ _____
Interpret Data 	What did you find out? What can you infer? _____ _____ What new question(s) does this raise? _____ _____ _____

Ask Questions 	What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	What code will make the table or display you want? _____ _____
Interpret Data 	What did you find out? What can you infer? _____ _____ What new question(s) does this raise? _____ _____ _____

Measures of Center and Spread

There are three ways to measure the **center** of a dataset, to summarize a whole column of quantitative data using just one number:

- The **mean** of a dataset is the average of all the numbers.
- The **median** of a dataset is a value that is smaller than half the dataset, and larger than the other half. In an ordered list the median will either be the middle number or the average of the two middle numbers.
- The **mode(s)** of a dataset is the value (or values) occurring most often. When all of the values occur equally often, a dataset has no mode.

In a **symmetric** dataset, values are just as likely to occur a certain distance above the mean as below the mean, and the median and mean are usually close together.

When a dataset is asymmetric, the median is a more descriptive measure of center than the mean.

- A dataset with **left skew** has a few values that are unusually low, which pull the mean *below* the median.
- A dataset with **right skew** has a few values that are unusually high, which pull the mean *above* the median.

When a dataset contains a small number of values, the mode may be the most descriptive measure of center.

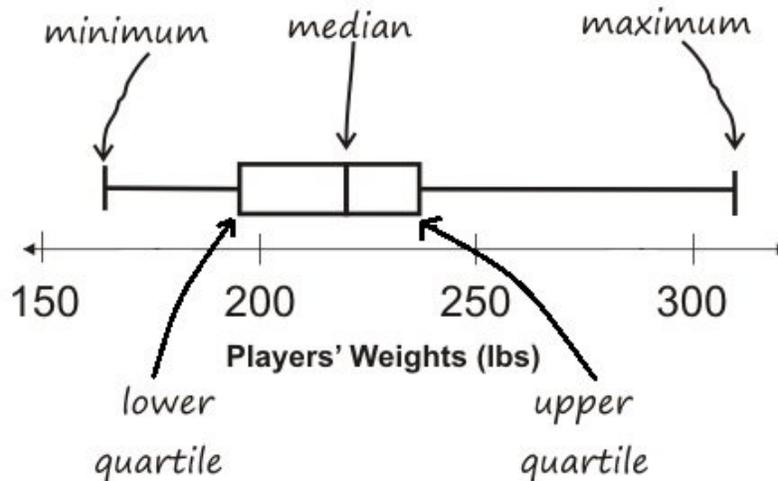
Data Scientists can also measure the **spread** of a dataset using a **five-number summary**:

- The **minimum** – the lowest value in the dataset
- The **first, or “lower” quartile (Q1)** – the middle of the lower half of values, which separates the lowest quarter from the next smallest quarter
- The **second quartile (Q2)** – the middle value, which separates the entire dataset into “top” and “bottom” halves
- The **third, or “upper” quartile (Q3)** – the middle of the higher half of values which separates the second highest quarter from the highest quarter
- The **maximum** – the largest value in the dataset

Measures of Center and Spread (continued)

The **five-number summary** can be used to draw a **box plot**.

- Each of the four sections of the box plot contains 25% of the data. *If the values are distributed evenly across the range, the four sections of the box plot will be equal in width.* Uneven distributions will show up as differently-sized sections of a box plot.
- The left **whisker** extends from the minimum to Q1.
- The **box**, or **interquartile range**, extends from Q1 to Q3. It is divided into 2 parts by the **median**. Each of those parts contains 25% of the data, so the whole box contains the central 50% of the data.
- The right **whisker** extends from Q3 to the maximum.



The box plot above, for example, tells us that:

- The minimum weight is about 165 pounds. The median weight is about 220 pounds. The maximum weight is about 310 pounds.
 - 1/4 of the players weigh roughly between 165 and 195 pounds
 - 1/4 of the players weigh roughly between 195 and 220 pounds
 - 1/4 of the players weigh roughly between 220 and 235 pounds
 - 1/4 of the players weigh roughly between 235 and 310 pounds
 - 50% of the players weigh roughly between 165 and 220 pounds
 - 50% of the players weigh roughly between 195 and 235 pounds
 - 50% of the players weigh roughly between 220 and 310 pounds
- The densest concentration of players' weights is between 220 and 235 pounds.
- Because the widest section of the box plot is between 235 and 310 pounds, we understand that the weights of the heaviest 25% fall across a wider span than the others. 310 may be an outlier, the weights of the players weighing between 235 pounds and 310 pound could be evenly distributed across the range, or all of the players weighing over 235 pounds may weigh around 310 pounds.

Summarizing Columns in the Animals Dataset

Find the measures of center and spread to summarize the _____ pounds _____ column of the Animals Table. Be sure to add examples to your Contracts page as you work.

Measures of Center

The three measures of center for this column are:

Mean (Average)	Median	Mode(s)

Since the mean is _____ compared to the median, this suggests the shape is _____.

higher/lower/about equal
skewed right (or high outliers) / skewed left (or low outliers) / symmetric

Measures of Spread

My five-number summary is:

Minimum	Q1	Median	Q3	Maximum

Displaying Center and Spread with a Box Plot

Draw a box plot from this summary on the number line below. Be sure to label the number line with consistent intervals.

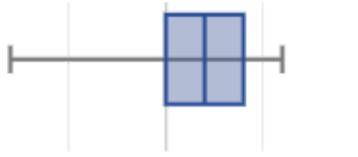


From this summary and box plot, I conclude:

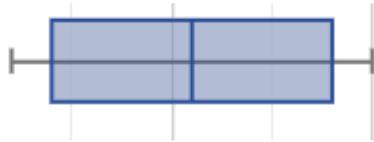
Identifying Shape - Box Plots

Describe the shape of the box plots on the left. Do your best to incorporate the vocabulary you've been introduced to.

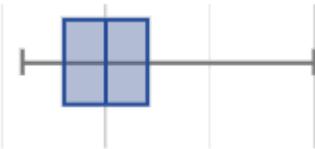
1



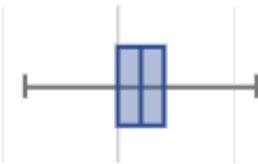
2



3



4

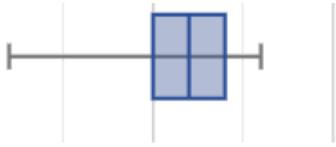


5



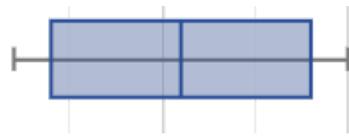
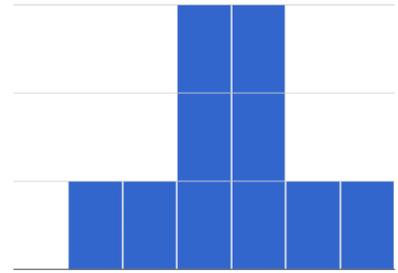
Matching Box Plots to Histograms

Students watched 5 videos, and rated them on a scale of 1 to 10. For each video, their ratings were used to generate box plots and histograms. Match the box plot to the histogram that displays the same data.



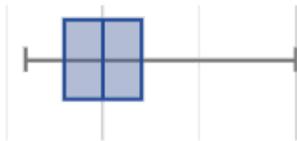
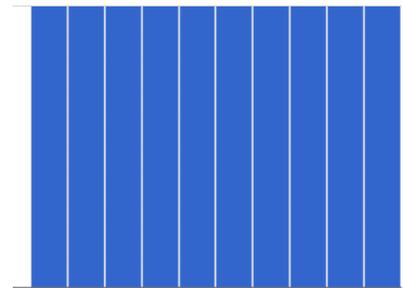
1

A



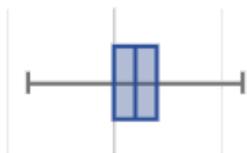
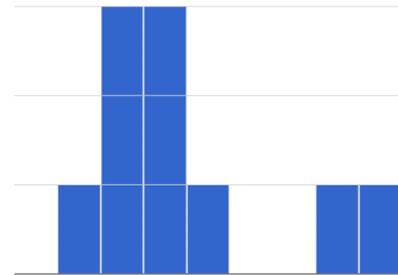
2

B



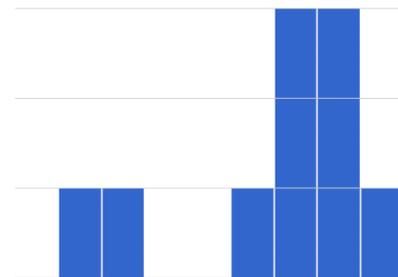
3

C



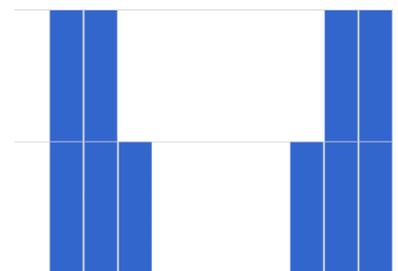
4

D



5

E



Data Cycle: Shape of My Dataset

Use the Data Cycle to explore the distribution of one or more quantitative columns using **histograms and box plots**, and write down your findings.

Ask Questions 	What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	What code will make the table or display you want? _____ _____
Interpret Data 	What did you find out? What can you infer? _____ _____ What new question(s) does this raise? _____ _____ _____

Ask Questions 	What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	What code will make the table or display you want? _____ _____
Interpret Data 	What did you find out? What can you infer? _____ _____ What new question(s) does this raise? _____ _____ _____

Shape of My Dataset

Find the measures of center and spread to summarize a column of your dataset.

The column I chose to summarize is: _____

Measures of Center

The three measures of center for this column are:

Mean (Average)	Median	Mode(s)

Since the mean is _____ compared to the median, this suggests the shape is

[higher/lower/about equal]

_____:

[skewed right (or high outliers) / skewed left (or low outliers) / symmetric]

Measures of Spread

My five-number summary is:

Minimum	Q1	Q2 (Median)	Q3	Maximum

Displaying Center and Spread with a Box Plot

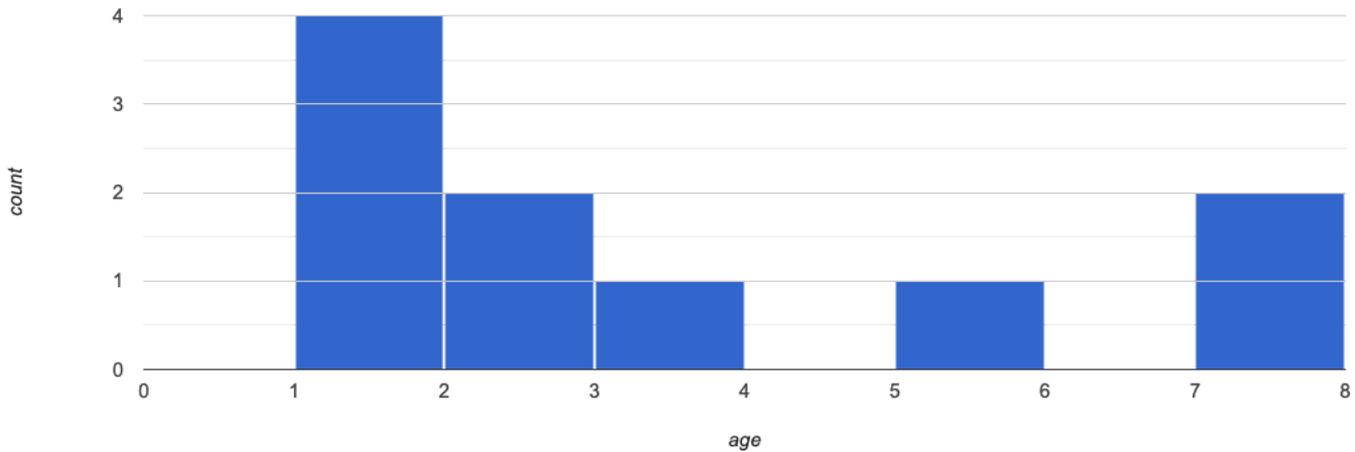
Draw a box plot from this summary on the number line below. Be sure to label the number line with consistent intervals.



From this summary and box plot, I conclude:

Computing Standard Deviation

The histogram below shows the ages of ten cats at the shelter:



- 1) Describe the shape of this histogram. _____
 2) How many cats are 1 year old? 2 years old? Fill in the table below by reading the histogram. The first column has been done for you.

age	1	2	3	4	5	6	7
count	4						

- 3) What is the mean age of the cats in this histogram? _____
 4) **Draw a star to locate the mean on the x-axis of the histogram above.**
 5) For each cat in the histogram above, **draw a horizontal arrow** under the axis from your star to the cat's interval, and **label the arrow with its distance from the mean.** (For example, if the mean is 3 and a cat is in the 1yr interval, your arrow would stretch from 1 to 3, and be labeled with the distance "2")
 6) We want to summarize all 10 distances from the mean into a single number. What number do you think will best summarize those distances?

To compute the standard deviation we square each distance and take the average, then take the square root of the average.

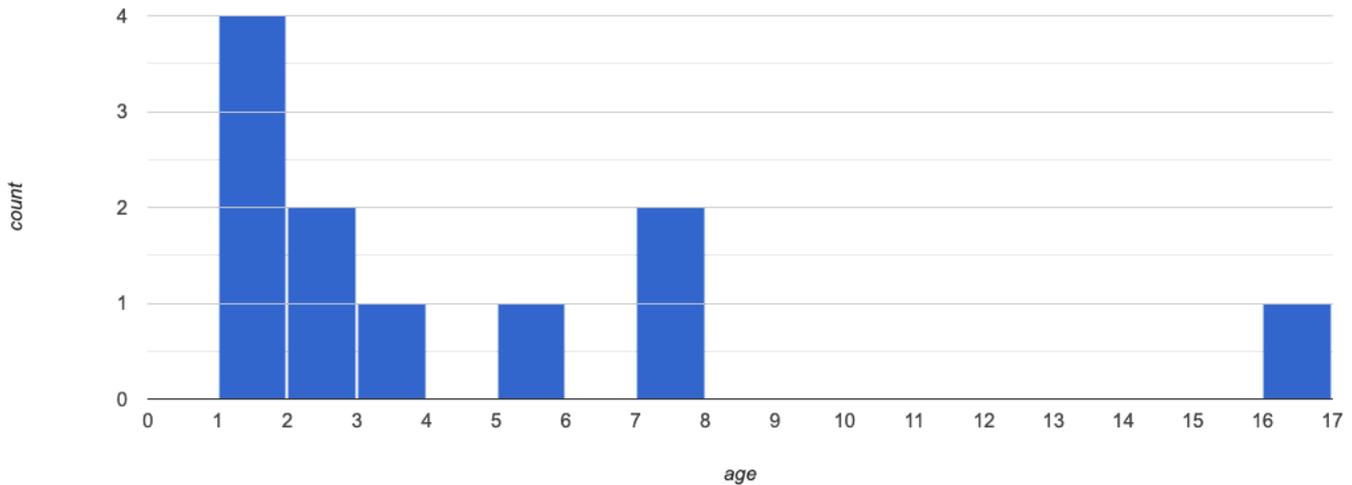
- 7) We've recorded the 10 ages shown in the histogram above in the table below and listed the distance-from-mean for the four 1-year-old cats for you. As you can see, 1 year-olds are 2 years away from the mean, so their squared distance is 4. Complete the table.

age of cat	1	1	1	1	2	2	3	5	7	7
distance from mean	2	2	2	2						
squared distance										

- 8) Add all the squared distances. What is their sum? _____
 9) What is the average squared distance? _____
 10) Take the square root to find the **standard deviation!** _____

The Effect of an Outlier

The histogram below shows the ages of eleven cats at the shelter:



1) Describe the shape of this histogram. _____

2) How many cats are 1 year old? 2 years old? Fill in the table below by reading the histogram. The first column has been done for you.

age	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
count	4															

3) What is the mean age of the cats in this histogram? _____

4) Draw a star to identify the mean on the histogram above.

5) For each cat in the histogram above, draw a horizontal arrow from the mean to the cat's interval, and label the arrow with its distance from the mean. (For example, if the mean is 2 and a cat is 5 years old, your arrow would stretch from 2 to 5, and be labeled with the distance "3")

To compute the standard deviation we square each distance and take the average, then take the square root of the average.

6) Recorded the 11 ages shown in the histogram in the first row of the table below. For each age, compute the distance from the mean and the squared distance.

age of cat																
distance from mean																
squared distance																

7) Add all the squared distances. What is their sum? _____

8) What is the average squared distance? _____

9) Take the square root to find the **standard deviation!** _____

10) How did the outlier impact the standard deviation? _____

Data Cycle: Standard Deviation in the Animals Dataset

The mean time-to-adoption is 5.75 weeks. Does that mean most animals generally get adopted in 4-6 weeks? Use the Data Cycle to find out. Write your findings on the lines below, in response to the question.

<p>Ask Questions</p> 	<p><i>Do the animals all get adopted in around the same length of time?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p> <hr/>
<p>Analyze Data</p> 	<p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p>What code will make the table or display you want?</p> <hr/>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/> <hr/>

Turn the Data Cycle above into a Data Story, which answers the question "If the average adoption time is 5.75 weeks, do all the animals get adopted in roughly 4-6 weeks?"

Data Cycle: Standard Deviation in My Dataset

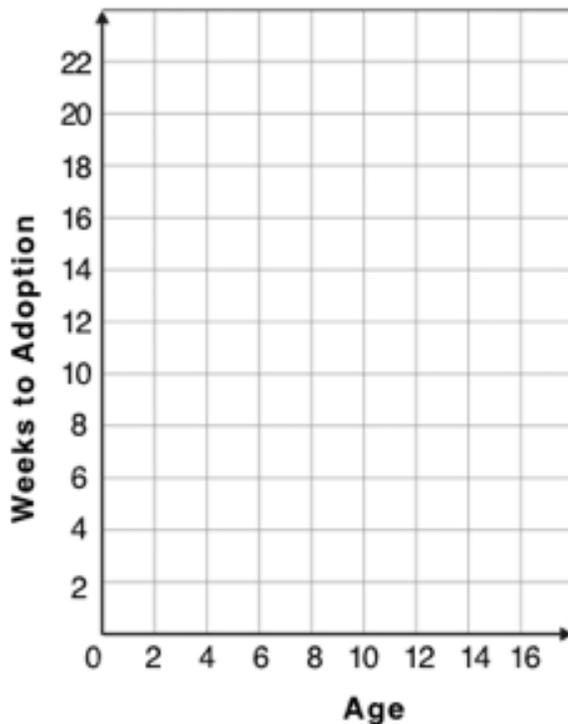
What questions do you have about the spread of your dataset? Use the Data Cycle to find the standard deviation in two distributions, and write down your thinking and findings.

Ask Questions 	What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	If you only need some rows, define your filter function here (Need help? Use the Design Recipe!) _____ If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!) _____ What code will make the table or display you want? _____
Interpret Data 	What did you find out? What can you infer? _____ _____ What new question(s) does this raise? _____ _____ _____

Ask Questions 	What question do you have? _____ _____ What type of question is this? (circle one) Lookup Arithmetic Statistical
Consider Data 	What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.) _____ What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.) _____
Analyze Data 	If you only need some rows, define your filter function here (Need help? Use the Design Recipe!) _____ If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!) _____ What code will make the table or display you want? _____
Interpret Data 	What did you find out? What can you infer? _____ _____ What new question(s) does this raise? _____ _____ _____

Creating a Scatter Plot

name	species	age	weeks
"Sasha"	"cat"	1	3
"Boo-boo"	"dog"	11	5
"Felix"	"cat"	16	4
"Buddy"	"lizard"	2	24
"Nori"	"dog"	6	9
"Wade"	"cat"	1	2
"Nibblet"	"rabbit"	6	12
"Maple"	"dog"	3	2

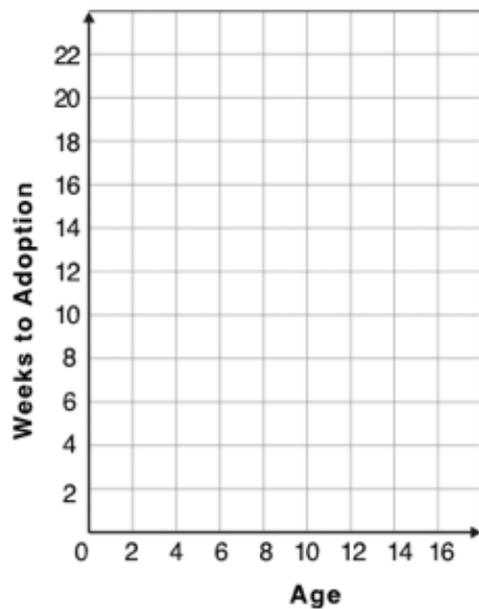


1) Choose an animal from the table on the left, and plot their age/weeks values by adding a dot to the scatter plot on the right (be sure to check your axes!). Then write their name next to your dot. Plot the rest of the animals - one at a time. After each one, ask yourself whether or not you

see a pattern in the data. After how many animals did you begin to see a pattern? _____
 Use a straight edge to draw a line on the graph that best represents the pattern you see, then circle the cloud of points around that line.

2) Are the points tightly clustered around the line or loosely scattered? _____
 3) Does this display support the claim that younger animals get adopted faster? Why or why not?

4 Place points on the graph to create a scatter plot with NO relationship.



Data Cycle: Relationships in the Animals Dataset

Use the Data Cycle to search for relationships! The first cycle has a question to get you started. What question will you ask for the second?

Ask Questions 	<p><i>Is there a relationship between weight and adoption time?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
Consider Data 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p> <hr/>
Analyze Data 	<p>What code will make the table or display you want?</p> <hr/>
Interpret Data 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/> <hr/>

Ask Questions 	<p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
Consider Data 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p> <hr/>
Analyze Data 	<p>What code will make the table or display you want?</p> <hr/>
Interpret Data 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/> <hr/>

Data Cycle: Relationships in Your Dataset

Use the Data Cycle to search for relationships between columns in your dataset!

Ask Questions 	<hr/> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
Consider Data 	<hr/> <p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>
Analyze Data 	<hr/> <p>What code will make the table or display you want?</p>
Interpret Data 	<hr/> <p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/> <hr/>

Ask Questions 	<hr/> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
Consider Data 	<hr/> <p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>
Analyze Data 	<hr/> <p>What code will make the table or display you want?</p>
Interpret Data 	<hr/> <p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/> <hr/>

Case Study: Ethics, Privacy, and Bias

These questions are designed to accompany one of the case studies provided in the [Ethics, Privacy, and Bias lesson](#).

My Case Study is _____

1) Read the case study you were assigned, and write your summary here.

2) Is this a good thing or a bad thing? Why?

3) What are the arguments on *each* side?

Data Science used for this purpose is good because...

Data Science used for this purpose is bad because...

Analyzing Survey Results When Data is Dirty

These questions are designed to accompany the [Survey of Eighth Graders and their Favorite Desserts Starter File](#).

1) Paolo made a **pie-chart** of the **dessert** column and was surprised to discover that **Fruit** was the most popular dessert among 8th graders! Make the **pie-chart** . Why is this pie-chart misleading? How is the data "dirty"?

2) What ideas do you have for how the survey designer could have made sure that the data in the **dessert** column would have been cleaner?

3) Shani made a **bar-chart** of the **gender-id** column. In her analysis she stated that the most common gender identity among eighth graders in her class is male. Make the **pie-chart** . Do you agree? Why or Why Not?

4) Make a chart showing the ages of the 8th graders surveyed. What "dirty" data problems do you spot and how are they misleading?

5) What ideas do you have for how the survey designer could have made sure that the data in the **age** column would have been cleaner?

Dirty Data!

A bunch of new animals are coming to the shelter, and that means more data! Open the [New Animals Dataset](#) and take a *careful* look.

What do you Notice?	What do you Wonder?

There are many different ways that data can be dirty!

- Missing Data** - A column containing some cells with data, but some cells left blank.
- Inconsistent Types** - A column with inconsistent data types. For example, a `years` column where almost every cell is a Number, but one cell contains the string `"5 years old"`.
- Inconsistent Units** - A column with consistent data types, but inconsistent units. For example, a `weight` column where some entries are in pounds but others are in kilograms.
- Inconsistent Naming** - Inconsistent spelling and capitalization for entries lead to them being counted as different. For example, a `species` column where some entries are `"cat"` and others are `"Cat"` will not give us a full picture of the cats.

1) Which animals' row(s) have **missing data**? _____

2) Which column(s) have **inconsistent types**? _____

3) Which column(s) have **inconsistent units**? _____

4) Which column(s) have **inconsistent naming**? _____

5) If we want to analyze this data, what should we do with the rows for Tanner, Toni, and Lizzy?

6) If we want to analyze this data, what should we do with the rows for Chanel and Bibbles?

7) If we want to analyze this data, what should we do with the rows for Porche and Boss?

8) If we want to analyze this data, what should we do with the row for Niko?

9) If we want to analyze this data, what should we do with rows for Mona, Rover, Susie Q, and Happy?

10) Sometimes data cleaning is straightforward. Sometimes the problem is evident but the solution is less certain. For which questions were you certain of your data cleaning suggestion? For which were you less certain? Why?

Bad Questions Make Dirty Data

The [Height vs. Wingspan Survey \(make a copy\)](#) has *lots* of problems, which can lead to many kinds of dirty data: Missing Data, Inconsistent Types, Inconsistent Units and Inconsistent Language! Try filling it out with bad data. Record the problems and make some recommendations for how to improve the survey!

NOTE: make sure your teachers gives you a link to their own copy of the survey. You should be able to the data after you submit it!

Ques	What examples of bad data were you able to submit?	How could the survey be improved to avoid bad data?
A		
B		
C		
D		

Looking up Rows and Columns

We can define names for values in Pyret, the same way we do in math:

```
name = "Flannery"  
age = 16  
logo = star(50, "solid", "red")
```

When **looking up a data Row** from a Table, programmers use the `row-n` method. This method takes a single number as its input, which tells the computer which Row we want. *Note: Rows are numbered starting at zero!*

For example:

```
sasha = animals-table.row-n(0) # define Sasha to be the first row  
mittens = animals-table.row-n(2) # define Mittens to be the third row
```

When we define these rows, it's more useful to name them based on their *properties*, rather than their identifiers:

```
cat-row = animals-table.row-n(0) # Sasha is a cat  
dog-row = animals-table.row-n(10) # Toggle is a dog
```

When **looking up a column** from a Row, programmers use square brackets and the name of the column they want.

For example:

```
# these two lines do the same thing! We can use the defined name to simplify our code  
animals-table.row-n(0)["age"] # look up Sasha's age (in row 0)  
cat-row["species"] # look up Sasha's age (using the defined name)  
dog-row["age"] # look up Toggle's age (using the defined name)
```

Lookup Questions

The table below represents four pets at an animal shelter:

`pets-table`

name	sex	age	pounds
"Toggle"	"female"	3	48
"Fritz"	"male"	4	92
"Nori"	"female"	6	35.3
"Maple"	"female"	3	51.6

1) Match each Lookup Question (left) to the code that will give the answer (right).

- | | | | |
|---------------------------------------|---|---|--|
| "How much does Maple weigh?" | 1 | A | <code>pets-table.row-n(3)</code> |
| "Which is the last row in the table?" | 2 | B | <code>pets-table.row-n(2)["name"]</code> |
| "What is Fritz's sex?" | 3 | C | <code>pets-table.row-n(1)["sex"]</code> |
| "What's the third animal's name?" | 4 | D | <code>pets-table.row-n(3)["age"]</code> |
| "How much does Nori weigh?" | 5 | E | <code>pets-table.row-n(3)["pounds"]</code> |
| "How old is Maple?" | 6 | F | <code>pets-table.row-n(0)</code> |
| "What is Toggle's sex?" | 7 | G | <code>pets-table.row-n(2)["pounds"]</code> |
| "What is the first row in the table?" | 8 | H | <code>pets-table.row-n(0)["sex"]</code> |

2) Write the Pyret code that will produce each value on the right.

a.	<u><code>pets-table.row-n(3)["name"]</code></u>	"Maple"
b.	<u>_____</u>	"male"
c.	<u>_____</u>	4
d.	<u>_____</u>	48
e.	<u>_____</u>	"Nori"

More Practice with Lookups

Consider the table below, and the four value definitions that follow:

shapes-table

name	corners	is-round
"triangle"	3	false
"square"	4	false
"rectangle"	4	false
"circle"	0	true

```
shapeA = shapes-table.row-n(0)
shapeB = shapes-table.row-n(1)
shapeC = shapes-table.row-n(2)
shapeD = shapes-table.row-n(3)
```

1) Match each Pyret expression (left) to the description of what it evaluates to (right).

- | | | | |
|---|---|---|---|
| <code>shapeD</code> | 1 | A | Evaluates to 4 |
| <code>shapeA</code> | 2 | B | Evaluates to the last row in the table |
| <code>shapeB["corners"]</code> | 3 | C | Evaluates to "square" |
| <code>shapeC["is-round"]</code> | 4 | D | Evaluates to true |
| <code>shapeB["name"]</code> | 5 | E | Evaluates to false |
| <code>shapeA["corners"]</code> | 6 | F | Evaluates to 3 |
| <code>shapeD["name"] == "circle"</code> | 7 | G | Evaluates to the first row in the table |

2) Fill in the blanks (left) with the Pyret lookup code that will produce the value (right).

a.	<hr/>	"rectangle"
b.	<hr/>	"square"
c.	<hr/>	4
d.	<hr/>	0
e.	<hr/>	true

Defining Rows

We've already given you two row definitions: `cat-row` and `dog-row` :

```
cat-row = animals-table.row-n(0) # Sasha is a cat
dog-row = animals-table.row-n(10) # Toggle is a dog
```

Remember: rows start at index zero! Use this to answer the questions below.

1 The index of a row containing a lizard is _____

2 The index of a row containing a rabbit is _____

3 The index of a row containing a fixed animal is _____

4 The index of a row containing a male animal _____

5 The index of a row containing a female animal is _____

6 The index of a row containing a hermaphroditic animal is _____

7 The index of a row containing an unfixed animal is _____

8 The index of a row containing a young animal (<2 years) is _____

9 The index of a row containing an old animal (>10 years) is _____

10) What code would you write to define `lizard-row` ?

11) What code would you write to define `rabbit-row` ?

12) What code would you write to define `fixed-row` ?

13) What code would you write to define `male-row` ?

14) What code would you write to define `female-row` ?

15) What code would you write to define `hermaphrodite-row` ?

16) What code would you write to define `young-row` ?

17) What code would you write to define `old-row` ?

Add this code to your Animals Starter File! You'll want these rows for later!

The Great gt domain debate!

Kermit: The domain of `gt` is `Number`, `String`, `String` .

Oscar: The domain of `gt` is `Number` .

Ernie: I'm not sure who's right!

In order to make a triangle, we need a size, a color and a fill style...

but all we had to tell our actor was `gt(20)` ...and they returned `triangle(20, "solid", "green")` .

Please help us!

1) What is the correct domain for `gt`?

2) What could you tell Ernie to help him understand how you know?

Let's Define Some New Functions!

1) Let's define a function `rs` to generate solid red squares of whatever size we give them!

If I say `rs(5)`, what would our actor need to say?

Let's write a few more examples:

`rs(_____)` → _____

`rs(_____)` → _____

`rs(_____)` → _____

What changes in these examples? Name your variable(s): _____

Let's define our function using the variable.

```
fun rs(_____): _____ end
```

2) Let's define a function `bigc` to generate big solid circles of size 100 in whatever color we give them!

If I say `bigc("orange")`, what would our actor need to say?

Let's write a few more examples:

`bigc(_____)` → _____

`bigc(_____)` → _____

`bigc(_____)` → _____

What changes in these examples? Name your variable(s): _____

Let's define our function using the variable.

```
fun bigc(_____): _____ end
```

3) Let's define a function `ps` to build a pink star of size 50, with the input determining whether it's solid or outline!

If I say `ps("outline")`, what would our actor need to say?

Write examples for all other possible inputs:

`ps(_____)` → _____

`ps(_____)` → _____

What changes in these examples? Name your variable(s): _____

Let's define our function using the variable.

```
fun ps(_____): _____ end
```

Add these new function definitions to your [gt Starter File](#) and test them out!

Let's Define Some More New Functions!

1) Let's define a function `sun` to write **SUNSHINE** in whatever color and size we give it!

If I say `sun(5, "blue")`, what would our actor need to say?

Let's write a few more examples:

`sun(_____, _____) → _____`

`sun(_____, _____) → _____`

`sun(_____, _____) → _____`

What changes in these examples? Name your variable(s): _____

Let's define our function using the variable.

```
fun sun(_____, _____): _____ end
```

2) Let's define a function `me` to generate your name in whatever size and color we give it!

If I say `me(18, "gold")`, what would our actor need to say?

Let's write a few more examples:

`me(_____, _____) → _____`

`me(_____, _____) → _____`

`me(_____, _____) → _____`

What changes in these examples? Name your variable(s): _____

Let's define our function using the variable.

```
fun me(_____, _____): _____ end
```

3) Let's define a function `gr` to build a solid, green rectangle of whatever length and width we give it!

If I say `gr(10, 80)`, what would our actor need to say?

Let's write a few more examples:

`gr(_____, _____) → rectangle(_____, _____, "solid", "green")`

`gr(_____, _____) → rectangle(_____, _____, "solid", "green")`

`gr(_____, _____) → rectangle(_____, _____, "solid", "green")`

What changes in these examples? Name your variable(s): _____

Let's define our function using the variable.

```
fun gr(_____, _____): _____ end
```

4) Add these new function definitions to your [gt Starter File](#) and test them out!

Describe and Define Your Own Functions!

1) Let's define a function _____ to generate...

If I say _____, what would our actor need to say? _____

Let's write a few more examples:

_____ (_____) → _____ (_____)

_____ (_____) → _____ (_____)

_____ (_____) → _____ (_____)

What variable changes? _____

Let's define our function using the variable.

```
fun _____ ( _____ ) : _____ ( _____ ) end
```

2) Let's define a function _____ to generate...

If I say _____, what would our actor need to say? _____

Let's write a few more examples:

_____ (_____) → _____ (_____)

_____ (_____) → _____ (_____)

_____ (_____) → _____ (_____)

What variable changes? _____

Let's define our function using the variable.

```
fun _____ ( _____ ) : _____ ( _____ ) end
```

3) Let's define a function _____ to generate...

If I say _____, what would our actor need to say? _____

Let's write a few more examples:

_____ (_____) → _____ (_____)

_____ (_____) → _____ (_____)

_____ (_____) → _____ (_____)

What variable changes? _____

Let's define our function using the variable.

```
fun _____ ( _____ ) : _____ ( _____ ) end
```

Add your new function definitions to your [gt Starter File](#) and test them out!

Defining Functions

Functions can be viewed in *multiple representations*. You already know one of them: Contracts, which specify the Name, Domain, and Range of a function. Contracts are a way of thinking of functions as a *mapping* between one set of data and another. For example, a mapping from Numbers to Strings:

```
f :: Number -> String
```

Another way to view functions is with Examples. Examples are essentially input-output tables, showing what the function would do for a specific input:

In our programming language, we focus on the last two columns and write them as code:

```
examples :  
  f(1) is 1 + 2  
  f(2) is 2 + 2  
  f(3) is 3 + 2  
  f(4) is 4 + 2  
end
```

Finally, we write a formal **function definition** ourselves. The pattern in the Examples becomes *abstract* (or "general"), replacing the inputs with variables. In the example below, the same definition is written in both math and code:

$$f(x) = x + 2$$

```
fun f(x): x + 2 end
```

Look for connections between these three representations!

- The function name is always the same, whether looking at the Contract, Examples, or Definition.
- The number of inputs in the Examples is always the same as the number of types in the Domain, which is always the same as the number of variables in the Definition.
- The "what the function does" pattern in the Examples is almost the same in the Definition, but with specific inputs replaced by variables.

Matching Examples and Contracts

Match each set of examples (left) with the contract that best describes it (right).

Examples

Contract

```
examples:  
f(5) is 5 / 2  
f(9) is 9 / 2  
f(24) is 24 / 2  
end
```

1

A

```
# f :: Number -> Number
```

```
examples:  
f(1) is rectangle(1, 1, "outline", "red")  
f(6) is rectangle(6, 6, "outline", "red")  
end
```

2

B

```
# f :: String -> Image
```

```
examples:  
f("pink", 5) is star(5, "solid", "pink")  
f("blue", 8) is star(8, "solid", "blue")  
end
```

3

C

```
# f :: Number -> Image
```

```
examples:  
f("Hi!") is text("Hi!", 50, "red")  
f("Ciao!") is text("Ciao!", 50, "red")  
end
```

4

D

```
# f :: Number, String -> Image
```

```
examples:  
f(5, "outline") is star(5, "outline", "yellow")  
f(5, "solid") is star(5, "solid", "yellow")  
end
```

5

E

```
# f :: String, Number -> Image
```

Matching Examples and Function Definitions

Highlight the variables in `gt` and label them with the word "size".

examples:

```
gt(20) is triangle(20, "solid", "green")
gt(50) is triangle(50, "solid", "green")
```

end

```
fun gt(size): triangle(size, "solid", "green") end
```

Highlight and label the variables in the example lists below. Then, using `gt` as a model, match the examples to their corresponding function definitions.

Examples			Definition
<pre>examples: f("solid") is circle(8, "solid", "red") f("outline") is circle(8, "outline", "red") end</pre>	1	A	<pre>fun f(s): star(s, "outline", "red") end</pre>
<pre>examples: f(2) is 2 + 2 f(4) is 4 + 4 f(5) is 5 + 5 end</pre>	2	B	<pre>fun f(num): num + num end</pre>
<pre>examples: f("red") is circle(7, "solid", "red") f("teal") is circle(7, "solid", "teal") end</pre>	3	C	<pre>fun f(c): star(9, "solid", c) end</pre>
<pre>examples: f("red") is star(9, "solid", "red") f("grey") is star(9, "solid", "grey") f("pink") is star(9, "solid", "pink") end</pre>	4	D	<pre>fun f(s): circle(8, s, "red") end</pre>
<pre>examples: f(3) is star(3, "outline", "red") f(8) is star(8, "outline", "red") end</pre>	5	E	<pre>fun f(c): circle(7, "solid", c) end</pre>

Creating Contracts From Examples

Write the contracts used to create each of the following collections of examples.

1) _____

examples:

```
big-triangle(100, "red") is triangle(100, "solid", "red")
big-triangle(200, "orange") is triangle(200, "solid", "orange")
end
```

2) _____

examples:

```
purple-square(15) is rectangle(15, 15, "outline", "purple")
purple-square(6) is rectangle(6, 6, "outline", "purple")
end
```

3) _____

examples:

```
banner("Game Today!") is text("Game Today!", 50, "red")
banner("Go Team!") is text("Go Team!", 50, "red")
banner("Exit") is text("Exit", 50, "red")
end
```

4) _____

examples:

```
twinkle("outline", "red") is star(5, "outline", "red")
twinkle("solid", "pink") is star(5, "solid", "pink")
twinkle("outline", "grey") is star(5, "outline", "grey")
end
```

5) _____

examples:

```
half(5) is 5 / 2
half(8) is 8 / 2
half(900) is 900 / 2
end
```

Contracts, Examples & Definitions - bc

gt

Directions: Define a function called `gt`, which makes solid green triangles of whatever size we want.

Every contract has three parts...

gt:: Number -> Image
function name Domain Range

Write some examples, then circle and label what changes...

examples:

gt(10) is triangle(10, "solid", "green")
function name input(s) what the function produces

gt(20) is triangle(20, "solid", "green")
function name input(s) what the function produces

end

Write the definition, giving variable names to all your input values...

fun gt(size):
function name variable(s)

triangle(size, "solid", "green")
what the function does with those variable(s)

end

bc

Directions: Define a function called `bc`, which makes solid blue circles of whatever radius we want.

Every contract has three parts...

:: ->
function name Domain Range

Write some examples, then circle and label what changes...

examples:

() is
function name input(s) what the function produces

() is
function name input(s) what the function produces

end

Write the definition, giving variable names to all your input values...

fun ():
function name variable(s)

what the function does with those variable(s)

end

Contracts, Examples & Definitions - Stars

sticker

Directions: Define a function called `sticker`, which consumes a color and draws a 50px star of the given color.

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)
_____ what the function does with those variable(s)

end

gold-star

Directions: Define a function called `gold-star`, which takes in a number and draws a solid gold star of that given size.

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)
_____ what the function does with those variable(s)

end

Contracts, Examples & Definitions - Name

name-color

Directions: Define a function called `name-color`, which makes an image of your name at size 50 in whatever color is given.

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)
_____ what the function does with those variable(s)

end

name-size

Directions: Define a function called `name-size`, which makes an image of your name in your favorite color (be sure to specify your name and favorite color!) in whatever size is given.

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)
_____ what the function does with those variable(s)

end

Defining Row-Consuming Functions

Suppose we wanted to label each data point on a scatter plot with its species...

1) Write the code to define `sample1` to be the 10th row in the table.

2) Write the code to look up the `species` column from `sample1`?

3) Write the code that would use the `text` function to show the species of `sample1` in red, 15px letters?

Instead of writing this out over and over for each animal, let's define a function to do it for us!

Species Tag

Directions: Define a function called `species-tag`, which takes in a row from the Animals Table and draws its name in red, 15px letters.

HINT: Use `cat-row` and `dog-row` in your examples!

Every contract has three parts...

`species-tag`:: _____ -> _____
function name Domain Range

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Write the definition, giving variable names to all your input values...

fun `species-tag`(_____): _____
function name variable(s)

_____ what the function does with those variable(s)

end

Custom Scatter Plot - Explore

Make sure you have [Custom Scatter Plot Starter File](#) open on your computer. Save a copy, and click "Run" to see the scatter-plot.

1) Comment-out the `scatter-plot` expression on line 53 by putting a `#` at the start of line. This will "turn off" the expression, by telling Pyret to ignore that line.

2) What is `species-tag` ? _____

3) Click "Run". What happens when you evaluate `species-tag(cat-row)` ?

4) What do you expect the `image-scatter-plot` expression on line 54 to produce?

5) Un-comment line 54 to "turn on" that code, then click "Run". What image appears?

6) What can you learn from this new display that you could not learn from the original image?

7) What ideas do you have for how else we could draw the dots in a scatter plot of the animals, so that it would be even more useful?

8) Find the other two "Helper Functions" in the file (`legs-tag` and `age-tag`). How are their definitions similar to `species-tag` ?

9) How are their definitions *different* from `species-tag` ?

10) Make a custom scatter plot using `legs-tag` . Then make another using `age-tag` . What is most interesting to you about the plots you just made?

11) Suppose we wanted to make a custom scatter plot where each point is labeled with the *name* of the animal. What would we need to define, to use in our custom scatter plot expression?

12) Make the new custom scatter plot described above. Write the definition of your helper function, and the custom scatter plot expression here:

species-dot - Explore

Look at the code in the [Custom Scatter Plot with Piecewise Functions Starter File](#).

What do you Notice?	What do you Wonder?

1) How is this program similar to the last one we looked at?

2) How is this code different from the program that made the `image-scatter-plot` with species labels?

3) Click "Run" and look at the display.

What do you Notice?	What do you Wonder?

Word Problem: sex-dot

Directions: We want to generate a custom dot for our `image-scatter-plot`, to distinguish female, hermaphrodite, and male animals. Write a function called `sex-dot`, which takes in a Row from the animals table and generates a 5-pixel solid dot that is red if the animal is female, purple if the animal is hermaphrodite and black if the animal is male.

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

_____ what the function does with those variable(s)

_____ what the function does with those variable(s)

_____ what the function does with those variable(s)

end

Defining Row Functions & Using Table Methods

Methods are special functions that are attached to pieces of data. We use them to manipulate Tables.

- In this course, the methods we'll be using are
 - `row-n` - consumes an index (starting with zero!) and produces a row from a table
 - `order-by` - consumes the name of a column and a Boolean value to determine if that table should be sorted by that column in ascending order
 - `filter` - consumes a *Boolean-producing function*, and produces a table containing only rows for which the function returns `true`
 - `build-column` - consumes the name of a new column, and a function that produces the values in that column for each Row
- Unlike functions, methods can't be used alone. They have a "secret" argument, which is the data they are attached to. They are written as part of that data, separated by a dot. For example:
`shapes.row-n(2)`
- Contracts for methods are different from other functions. They include the type of the data as part of their names. For example:
`<table>.row-n :: (index :: Number) -> Row`

Reading Row and Function Definitions

Make sure you've opened the [Table Methods Starter File](#) on your computer.

- 1) What name is being defined on line 8? _____
- 2) How many columns are listed? _____
- 3) What name is being defined on line 15? _____
- 4) Is `cat-row` a Number, String, Image or Row? _____
- 5) Type `cat-row` into the Interactions Area. What do you get? _____
- 6) A Contract for a function is written on line 32. What is its name? _____
- 7) What is its Domain? _____
- 8) What is its Range? _____
- 9) All 3 of the examples on lines 38-40 do the same thing, at different levels of abstraction. Think about what each example is doing and write a similar example using `dog-row`

Examples with <code>cat-row</code>	Similar examples with <code>dog-row</code>
<code>is-dog(cat-row) is false</code>	
What is this example doing?	
<code>is-dog(cat-row) is "cat" == "dog"</code>	
What is this example doing?	
<code>is-dog(cat-row) is cat-row["species"] == "dog"</code>	
What is this example doing?	

- 10) What other functions are defined here? _____
- 11) Evaluate `is-dog(dog-row)`. What do you get? _____
- 12) Evaluate `is-dog(cat-row)`. What do you get? _____
- 13) What does `is-old` do? _____
- 14) What does `lookup-name` do? _____
- 15) What does `lookup-fixed` do? _____
- 16) What does `label` do? _____
- 17) What does `kilos` do? _____

Exploring Methods

Make sure you've opened the [Table Methods Starter File](#) on your computer, and clicked "Run".

Ordering Rows

1) What does `animals-table.order-by("name", true)` evaluate to?

2) What does `animals-table.order-by("age", false)` evaluate to?

3) Sort the animals table from heaviest to lightest.

4) Sort the animals table alphabetically by *species*.

5) Sort the animals table by how long it took for each animal to be adopted, from longest to shortest adoption time.

6) What do `.order-by` and `.row-n` have in common? How are they different?

Filtering Rows

7) What does `animals-table.filter(lookup-fixed)` evaluate to?

8) What does `animals-table.filter(is-dog)` evaluate to?

9) Write the code to generate a table showing only the old animals. _____

10) What do the contracts for `lookup-fixed`, `is-dog`, and `is-old` have in common?

11) What happens if you evaluate `animals-table.filter(lookup-name)`? _____

12) Find the Contract for `.filter`. What is its Domain? _____

13) Where else have we seen functions that consume functions? _____

Building Columns

14) What does `animals-table.build-column("old", is-old)` evaluate to?

15) What does `animals-table.build-column("sticker", label)` evaluate to?

16) In your own words, how does the `.build-column` method work?

What Table Do We Get?

You have the following functions defined below :

```
fun lookup-fixed(animal): animal["fixed"]           end
fun is-dog(animal):      animal["species"] == "dog"  end
fun is-old(animal):      animal["age"] > 10          end
fun label(animal):       text(animal["name"], 20, "red") end
```

The table `t` below represents four animals from the shelter:

name	sex	age	fixed	species	pounds
"Toggle"	"female"	12	true	"dog"	48
"Fritz"	"male"	4	false	"dog"	92
"Nori"	"female"	6	true	"dog"	35.3
"Sunflower"	"female"	2	false	cat	51.6

Below is a list of expressions that use table methods. Match each expression to the description of the table we should get after evaluating each one.

`t.order-by("age", true)` 1

A Produces a table with Toggle, Fritz, and Nori - but not Sunflower.

`t.order-by("pounds", false)` 2

B Produces a table of all four animals, sorted youngest-to-oldest

`t.build-column("sticker", label)` 3

C Produces a table, with only Toggle.

`t.filter(is-old)` 4

D Produces an identical table with an extra column called "dog", whose values are true, true, true, false

`t.filter(lookup-fixed)` 5

E Produces a table containing only Nori and Toggle.

`t.filter(is-dog)` 6

F Produces a table with all four animals, sorted from heaviest to lightest.

`t.build-column("dog", is-dog)` 7

G Won't run: will produce an error

`t.filter(label)` 8

H Produces an identical table with an extra column called "sticker", whose values are images

Solving Word Problems

Being able to see functions as Contracts, Examples or Definitions is like having three powerful tools. These representations can be used together to solve word problems!

- 1) When reading a word problem, the first step is to figure out the **Contract** for the function you want to build. Remember, a Contract must include the Name, Domain and Range for the function!
- 2) Then we write a **Purpose Statement**, which is a short note that tells us what the function *should do*. Professional programmers work hard to write good purpose statements, so that other people can understand the code they wrote!
- 3) Next, we write at least two **Examples**. These are lines of code that show what the function should do for a *specific* input. Once we see examples of at least two inputs, we can *find a pattern* and see which parts are changing and which parts aren't.
- 4) To finish the Examples, we circle the parts that are changing, and label them with a short **variable name** that explains what they do.
- 5) Finally, we **define the function** itself! This is pretty easy after you have some examples to work from: we copy everything that didn't change, and replace the changeable stuff with the variable name!

Matching Word Problems and Purpose Statements

Match each word problem below to its corresponding purpose statement.

Annie got a new dog, Xavier, that eats about 5 times as much as her little dog, Rex, who is 10 years old. She hasn't gotten used to buying enough dogfood for the household yet. Write a function that generates an estimate for how many pounds of food Xavier will eat, given the amount of food that Rex usually consumes in the same amount of time.

1

A

Consume the pounds of food Rex eats and add 5.

Adrienne's raccoon, Rex, eats 5 more pounds of food each week than her pet squirrel, Lili, who is 7 years older. Write a function to determine how much Lili eats in a week, given how much Rex eats.

2

B

Consume the pounds of food Rex eats and subtract 5.

Alejandro's rabbit, Rex, poops about $\frac{1}{5}$ of what it eats. His rabbit hutch is 10 cubic feet. Write a function to figure out how much rabbit poop Alejandro will have to clean up depending on how much Rex has eaten.

3

C

Consume the pounds of food Rex eats and multiply by 5.

Max's turtle, Rex, eats 5 pounds less per week than his turtle, Harry, who is 2 inches taller. Write a function to calculate how much food Harry eats, given the weight of Rex's food.

4

D

Consume the pounds of food Rex eats and divide by 5.

Writing Examples from Purpose Statements

We've provided contracts and purpose statements to describe two different functions. Write examples for each of those functions.

Contract and Purpose Statement

Every contract has three parts...

triple:: _____ *Number* _____ -> *Number*
function name Domain Range

Consumes a Number and triples it.
what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Contract and Purpose Statement

Every contract has three parts...

upside-down:: _____ *Image* _____ -> *Image*
function name Domain Range

Consumes an image, and turns it upside down by rotating it 180 degrees.
what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Fixing Purpose Statements

Beneath each of the word problems below is a purpose statement that is either missing information or includes unnecessary information. Write an improved version of each purpose statement beneath the original.

1) **Word Problem:** *The New York City ferry costs \$2.75 per ride. The Earth School requires two chaperones for any field trip. Write a function `fare` that takes in the number of students in the class and returns the total fare for the students and chaperones.*

Purpose Statement: Define a function `fare` to take in the number of students and add 2.

Improved Purpose Statement:

2) **Word Problem:** *It is tradition for the Green Machines to go to Humpy Dumpty's for ice cream with their families after their soccer games. Write a function `cones` to take in the number of kids and calculate the total bill for the team, assuming that each kid brings two family members and cones cost \$1.25.*

Purpose Statement: Define a function `cones` to take in the number of kids on the team and multiply it by 1.25.

Improved Purpose Statement:

3) **Word Problem:** *The cost of renting an ebike is \$3 plus an additional \$0.12 per minute. Write a function `ebike` that will calculate the cost of a ride, given the number of minutes ridden.*

Purpose Statement: Define a function `ebike` to take in the number of minutes and multiply it by 3.12.

Improved Purpose Statement:

4) **Word Problem:** *Suleika is a skilled house painter at only age 21. She has painted hundreds of rooms and can paint about 175 square feet an hour. Write a function `paint` that takes in the number of square feet of the job and calculates how many hours it will take her.*

Purpose Statement: Define a function `paint` to take in the number of square feet of walls in a house and divide them by 175 to calculate the number of hours that it will take 21 year-old Suleika to complete the paint job.

Improved Purpose Statement:

Word Problem: rocket-height

Directions: A rocket blasts off, and is now traveling at a constant velocity of 7 meters per second. Use the Design Recipe to write a function `rocket-height`, which takes in a number of seconds and calculates the height.

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

_____ what the function does with those variable(s)

end

Method Chaining

Method chaining allows us to apply multiple methods with less code.

For example, instead of using multiple definitions, like this:

```
with-labels = animals-table.build-column("labels", nametag)
cats = with-labels.filter(is-cat)
cats.order-by("age", true)
```

We can use method-chaining to write it all on one line, like this:

```
animals-table.build-column("labels", nametag).filter(is-cat).order-by("age", true)
```

Order Matters! The methods are applied in the order they appear. For example, trying to order a table by a column that hasn't been built will result in an error.

The Design Recipe: is-dog / is-female

Directions: Define a function called `is-dog`, which consumes a `Row` of the animals table and *computes* whether the animal is a dog. HINT: use predefined rows like `dog-row` to make your examples easier!

Contract and Purpose Statement

Every contract has three parts...

`is-dog`:: Row \rightarrow Boolean
function name Domain Range

Consumes an animal, and checks whether the species == "dog"
what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

`is-dog`(`dog-row`) is `dog-row["species"] == "dog"`
function name input(s) what the function produces

`is-dog`() is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

what the function does with those variable(s)

end

Directions: Define a function called `is-female`, which consumes a `Row` of the animals table and returns true if the animal is female. HINT: use predefined rows like `female-row` to make your examples easier!

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ \rightarrow _____
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

what the function does with those variable(s)

end

Chaining Methods

You have the following functions defined below (read them *carefully!*):

```
fun is-fixed(r): r["fixed"]           end
fun is-young(r): r["age"] < 4       end
fun nametag(r):  text(r["name"], 20, "red") end
```

The table `t` below represents four animals from the shelter:

name	sex	age	fixed	pounds
"Toggle"	"female"	3	true	48
"Fritz"	"male"	4	true	92
"Nori"	"female"	6	true	35.3
"Maple"	"female"	3	true	51.6

Match each Pyret expression (left) to the description of what it does (right).

- | | | | |
|--|---|---|--|
| <code>t.order-by("age", true)</code> | 1 | A | Produces a table containing only Toggle and Maple |
| <code>t.filter(is-fixed)</code> | 2 | B | Produces a table of only young, fixed animals |
| <code>t.build-column("sticker", nametag)</code> | 3 | C | Produces a table, sorted youngest-to-oldest |
| <code>t.filter(is-young)</code> | 4 | D | Produces a table with an extra column, named "sticker" |
| <code>t.filter(is-young)
.filter(is-fixed)</code> | 5 | E | Produces a table containing Maple and Toggle, in that order |
| <code>t.filter(is-young)
.order-by("pounds", false)</code> | 6 | F | Produces a table containing the same four animals |
| <code>t.build-column("label", nametag)
.order-by("age", true)</code> | 7 | G | Won't run: will produce an error |
| <code>t.order-by("sx", false)</code> | 8 | H | Produces a table with an extra "label" column, sorted youngest-to-oldest |

Chaining Methods 2: Order Matters

You have the following functions defined below (read them *carefully!*):

```
fun is-female(r): r["sex"] == "female" end
fun kilograms(r): r["pounds"] / 2.2 end
fun is-heavy(r): r["kilos"] > 25 end
```

The table `t` below represents four animals from the shelter:

name	sex	age	fixed	pounds
"Toggle"	"female"	3	true	48
"Fritz"	"male"	4	true	92
"Nori"	"female"	6	true	35.3
"Maple"	"female"	3	true	51.6

Match each Pyret expression (left) to the description of what it does (right). **Note: one description might match multiple expressions!**

- | | | | |
|--|---|---|--|
| <code>t.order-by("kilos", true)</code> | 1 | A | Produces a table containing Toggle, Nori and Maple, with an extra column showing their weight in kilograms |
| <code>t.filter(is-female)
.build-column("kilos", kilograms)</code> | 2 | B | Produces a table containing Maple, Nori and Toggle (in that order) |
| <code>t.build-column("kilos", kilograms)
.filter(is-heavy)</code> | 3 | C | Produces a table containing only Fritz. |
| <code>t.filter(is-heavy)
.build-column("kilos", kilograms)</code> | 4 | D | Won't run: will produce an error (if so, why?) |
| <code>t.build-column("kilos", kilograms)
.filter(is-heavy)
.order-by("sex", true)</code> | 5 | E | Produces a table containing only Fritz, with two extra columns. |
| <code>t.build-column("female", is-female)
.build-column("kilos", kilograms)
.filter(is-heavy)</code> | 6 | F | Produces a table containing Maple and Fritz |

Defining Table Functions

The steps of the Design Recipe don't change just because we're working with Rows, but we can make some adjustments when using Row-consuming functions to filter tables and build columns! Let's try a concrete example:

Write a function `is-lizard`, which tells us whether an animal is a lizard.

Contract and Purpose

- We still want to pick good names. Are we writing a function to check if an animal is a lizard? Call it `is-lizard` !
- The Domain is a lot easier – it's *always* a `Row` !
- The Range is easier, too. If we're writing a function to filter a Table, we know the Range *has to be a Boolean* . (What would it be if we were building a column of Numbers? Images? Strings?)

Examples

The goal of the Examples step is to *find the pattern* that represents what the function does. When working with Rows, sometimes we have to start by just focusing on what the answer should be.

Suppose `animalC` is a lizard, and `animalD` is a cat. We can imagine the answers for an `is-lizard` to be...

```
examples:  
  is-lizard(animalC) is true  
  is-lizard(animalD) is false  
end
```

But how do we know these are `true` and `false` ? Well, we KNOW `animalC` is a lizard, and we KNOW `animalD` is a cat. So let's replace those answers with the Boolean expressions that compare their species:

```
examples:  
  is-lizard(animalC) is "lizard" == "lizard" # will produce true  
  is-lizard(animalD) is "cat" == "lizard"    # will produce false  
end
```

But what work gives us "lizard" and "cat"? Well, we can look in the `species` column!

```
examples:  
  is-lizard(animalC) is animalC["species"] == "lizard" # will produce true  
  is-lizard(animalD) is animalD["species"] == "lizard" # will produce false  
end
```

Sometimes we can go straight to this final form, doing the whole thing in one step. But it's nice to know we can break it down into pieces if we have to.

Once we see the pattern, we can *circle and label what changes* . In this case, only the Row representing the animal changes! We might use `r` as label, to represent the Row.

Definition

The final step in the Design Recipe is to take the pattern from our examples and *generalize it* to work with any input. It's no different when working with Rows.

Once again, our previous step is a huge help. We can simply **copy everything that stays the same**, and replace the part that changes with the label we used:

```
fun is-lizard(r): r["species"] == "lizard" end
```

Making Connections

This workbook page references the [Row Functions Starter File](#). Open that file on your computer, save a copy, and Click "Run"!

1) Find the definition for `cat-row`. What do you get back when you evaluate `cat-row` in the Interactions Area? _____

2) How many weeks did it take for `cat-row` to be adopted? _____ How many for `young-row`? _____

3) Scroll down in the starter file until you find the Contract, Purpose, Examples and Definition for `weeks-dot`. Discuss with your partner:

- What is the Domain of this function? _____ The Range? _____
- The Purpose Statement is a way of describing the function in detail. Does the Purpose Statement make it clear what this function should do, when given a Row?
- How many examples do you see defined for this function? _____

4) Look at the first two examples. How are they connected to the Contract and Purpose Statement?

These examples show us exactly what should be produced for `cat-row` and `young-row` - the two Rows representing "Miaulis" and "Nori", based on their `weeks` to adoption. But these examples only tell us part of the story! Where does the computer get the number of `weeks` from?

5) Now look at the *last* two examples. How are they connected to the first two?

6) Now look at the definition. How is this connected to our examples?

7) We've learned that representations of functions have to match.

Look at the Examples carefully - **there's a mistake**, where the Examples and Definition don't match the Contract and Purpose.

What's the mistake?

When you're done, fix the mistake in the code!

Data Cycle: Analyzing Dogs

<p>Ask Questions</p> 	<p><i>How many young dogs are fixed v. unfixed?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p> <hr/>
<p>Analyze Data</p> 	<p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p>What code will make the table or display you want?</p> <hr/>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/> <hr/>
<p>Ask Questions</p> 	<p><i>Is there a relationship between weight and days-to-adoption for dogs?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p> <hr/>
<p>Analyze Data</p> 	<p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p>What code will make the table or display you want?</p> <hr/>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/> <hr/>

Data Cycle: Analyzing Cats

<p>Ask Questions</p> 	<p><i>Are more old cats are fixed than unfixed unfixed?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p> <hr/>
<p>Analyze Data</p> 	<p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p>What code will make the table or display you want?</p> <hr/>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/>

<p>Ask Questions</p> 	<p><i>Is there a relationship between the number of kilograms that a cat weighs, and how long it takes to be adopted?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p> <hr/>
<p>Analyze Data</p> 	<p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p>What code will make the table or display you want?</p> <hr/>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/>

Design Recipe

Directions:

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

_____ what the function does with those variable(s)

end

Directions:

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

_____ what the function does with those variable(s)

end

Grouped Samples from the Animals Dataset

Use method chaining to define the **grouped samples** below. Assume that the following helper functions are defined exactly the way they are in the [Grouped Samples Starter File](#): `is-young`, `is-cat`, `is-dog`, `is-female`, `is-fixed`, and `name-has-s`. We've given you the solution for the first sample, to get you started.

	Subset	The code to define that subset
1	Kittens	<code>kittens = animals-table.filter(is-cat).filter(is-young)</code>
2	Puppies	
3	Fixed Cats	
4	Cats with "s" in their name	
5	Old Dogs	
6	Fixed Animals	
7	Old Female Cats	
8	Fixed Kittens	
9	Fixed Female Dogs	
10	Old Fixed Female Cats	

Displaying Data

Fill in the tables below, then use Pyret to make the following displays. Record the code you used. The first table has been filled in for you.

1) A `bar-chart` showing how many puppies are fixed or not.

What Rows?	Which Column(s)?	What will you Create?
<i>puppies</i>	<i>fixed</i>	<i>bar-chart</i>

code: `bar-chart(animals-table.filter(is-dog).filter(is-young), "fixed")`

2) A `pie-chart` showing how many heavy dogs are fixed or not.

What Rows?	Which Column(s)?	What will you Create?

code: _____

3) A `histogram` of the number of `weeks` it takes for a random sample of animals to be adopted.

What Rows?	Which Column(s)?	What will you Create?

code: _____

4) A `box-plot` of the number of `pounds` that kittens weigh.

What Rows?	Which Column(s)?	What will you Create?

code: _____

5) A `scatter-plot` of a random sample using `species` as the labels, `age` as the x-axis, and `weeks` as the y-axis.

What Rows?	Which Column(s)?	What will you Create?

code: _____

6) Describe **your own grouped sample** here, and fill in the table below.

What Rows?	Which Column(s)?	What will you Create?

code: _____

Data Cycle: Analyzing Categorical Data

<p>Ask Questions</p> 	<p><i>How many of each species are fixed at the shelter?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p> <hr/>
<p>Analyze Data</p> 	<p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p>What code will make the table or display you want?</p> <hr/>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/>

<p>Ask Questions</p> 	<p><i>Are there more female cats than male cats at the shelter?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p> <hr/>
<p>Analyze Data</p> 	<p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p>What code will make the table or display you want?</p> <hr/>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/>

Samples from My Dataset

Think back to when we defined grouped samples from the Animals Table, like "puppies", "old cats", etc. What grouped samples would be useful for *your* dataset? List a few of these in the first column.

Then, for each one, what function will identify if a row `r` is in the subset?

Grouped Sample	A function that returns true if a row <code>r</code> is in the subset
	<pre>fun _____(r): end</pre>

The Design Recipe

Write helper functions for **your** dataset, which you can use to define grouped samples. Since all helper functions will consume Rows, their Domains have already been filled in for you.

Directions: Define a function called _____, which consumes a Row of the _____ table and _____.

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ (*r* :: *Row*) _____ -> *Boolean*
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

_____ what the function does with those variable(s)

end

Directions: Define a function called _____, which consumes a Row of the _____ table and _____.

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ (*r* :: *Row*) _____ -> *Boolean*
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

_____ what the function does with those variable(s)

end

The Design Recipe

Write helper functions for **your** dataset, which you can use to define grouped samples. Since all helper functions will consume Rows, their Domains have already been filled in for you.

Directions: Define a function called _____, which consumes a Row of the _____ table and _____.

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ $(r :: Row)$ -> *Boolean*
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

what the function does with those variable(s)

end

Directions: Define a function called _____, which consumes a Row of the _____ table and _____.

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ $(r :: Row)$ -> *Boolean*
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

what the function does with those variable(s)

end

Correlations in Scatter Plots

Scatter Plots can be used to show a relationship between two quantitative columns. Each row in the dataset is represented by a point, with one column providing the x-value and the other providing the y-value.

The resulting “point cloud” makes it possible to look for a relationship between those two columns.

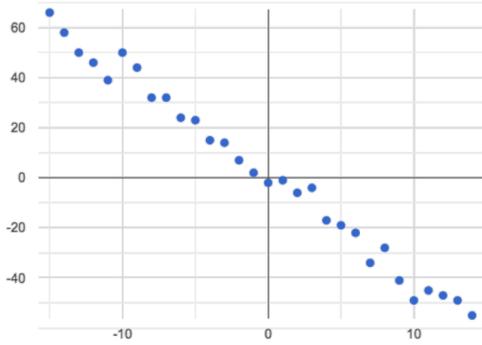
- If the points in a scatter plot appear to follow a straight line, it suggests that a linear relationship exists between those two columns. A number called a **correlation** can be used to summarize this relationship.
- r is the name of the **correlation statistic**. The r -value will always fall between -1 and $+1$. The sign tells us whether the correlation is positive or negative. Distance from 0 tells us the strength of the correlation.
 - -1 is the strongest possible negative correlation.
 - $+1$ is the strongest possible positive correlation.
 - 0 means no correlation.
- The correlation is **positive** if the point cloud slopes up as it goes farther to the right. This means larger y-values tend to go with larger x-values. The correlation is **negative** if the point cloud slopes down as it goes farther to the right.
- It is a **strong** correlation if the points are tightly clustered around a line. In this case, knowing the x-value gives us a pretty good idea of the y-value. It is a **weak** correlation if the points are loosely scattered and the y-value doesn't depend much on the x-value.
- Points that are far above or below the cloud of points in a scatter plot are called **outliers**.
- We graphically summarize this relationship by drawing a straight line through the data cloud, so that the vertical distance between the line and all the points taken together is as small as possible. This line is called the **line of best fit** and allows us to predict y-values based on x-values.
- **Correlation is not causation!** Correlation only suggests that two column variables are related, but does not tell us if one causes the other. For example, hot days are correlated with people running their air conditioners, but air conditioners do not cause hot days!

Identifying Form, Direction and Strength

Can you identify the Form, Direction, & Strength of these displays?

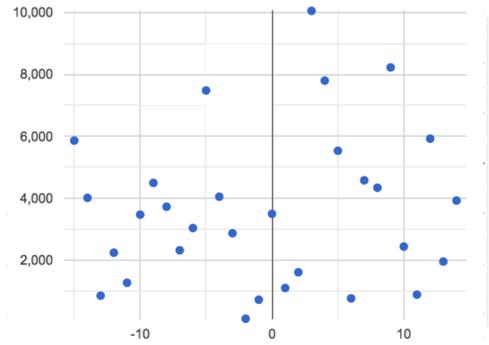
Note: If the form is non-linear, we shouldn't report direction - a curve may rise and then fall.

A



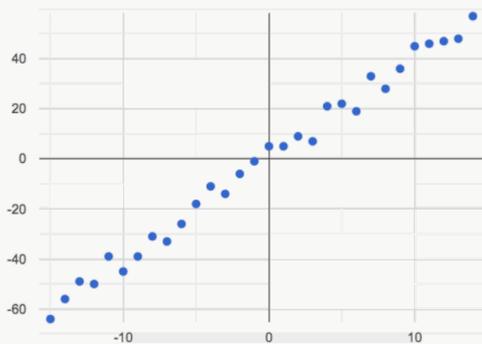
Form: Linear Non-Linear None
Direction: Positive Negative None
Strength: Strong Weak None

B



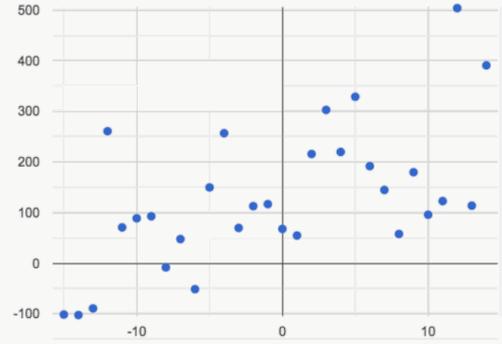
Form: Linear Non-Linear None
Direction: Positive Negative None
Strength: Strong Weak None

C



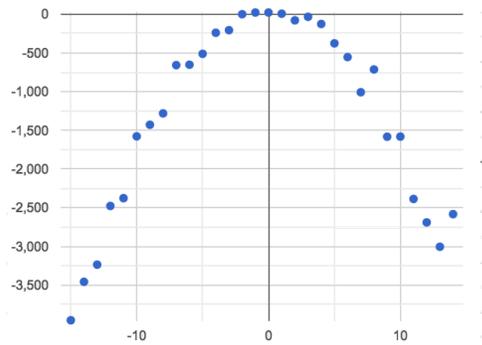
Form: Linear Non-Linear None
Direction: Positive Negative None
Strength: Strong Weak None

D



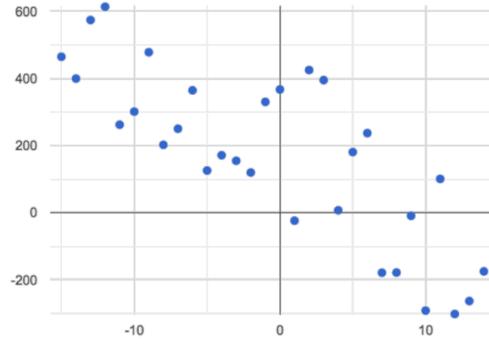
Form: Linear Non-Linear None
Direction: Positive Negative None
Strength: Strong Weak None

E



Form: Linear Non-Linear None
Direction: Positive Negative None
Strength: Strong Weak None

F



Form: Linear Non-Linear None
Direction: Positive Negative None
Strength: Strong Weak None

Reflection on Form, Direction and Strength

1) What has to be true about the *shape* of a relationship in order to start talking about a correlation?

2) What is the difference between a *weak* relationship and a *negative* relationship?

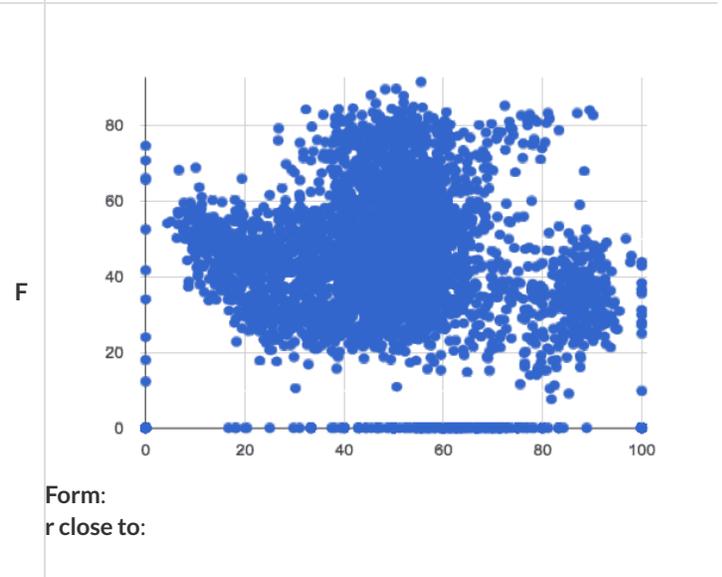
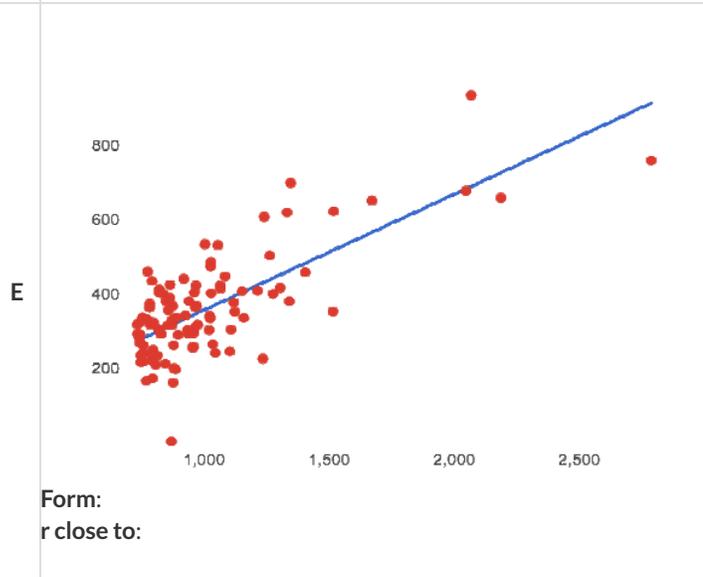
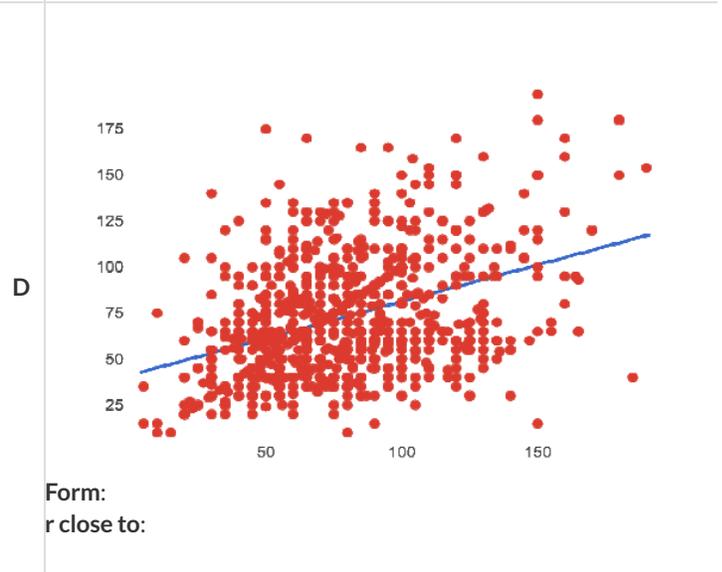
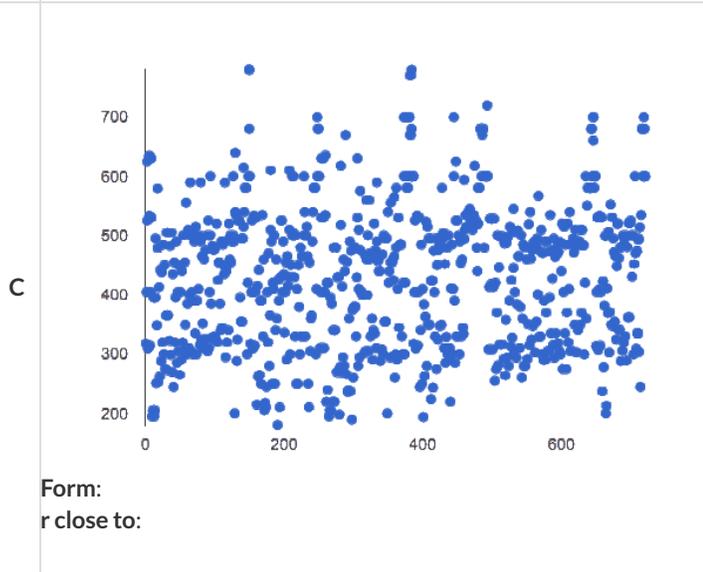
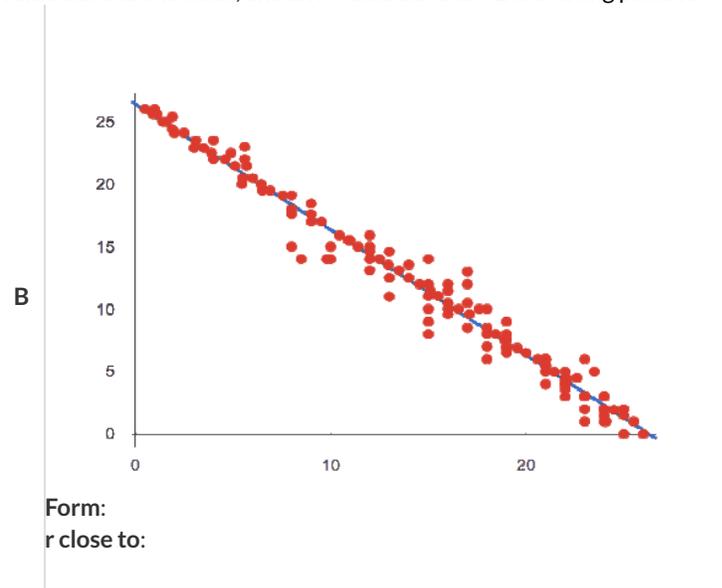
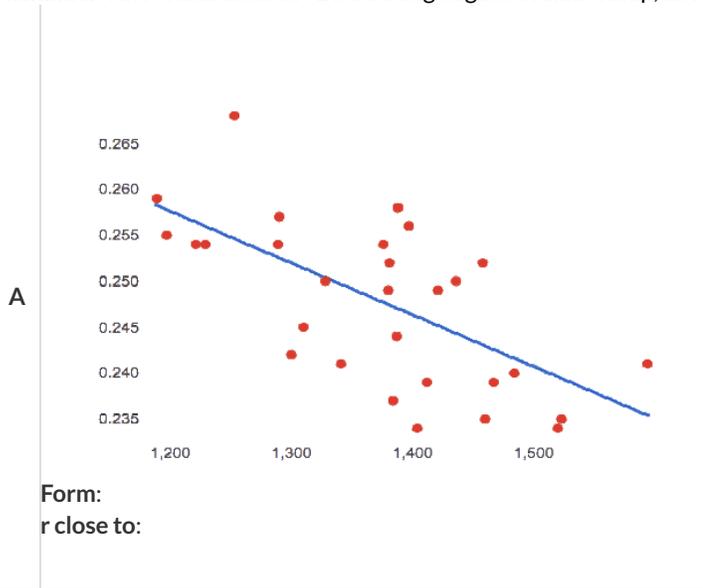
3) What is the difference between a *strong* relationship and a *positive* relationship?

4) If we find a strong relationship in a sample from a larger population, will that relationship *always hold* for the whole population? Why or why not?

5) A news report claims that after surveying *10 million people*, a positive correlation was found between how much chocolate a person eats and how happy they are. Does this mean eating chocolate almost certainly makes you happier? Why or why not?

Identifying Form and r-Values

Can you identify the Form and r-Values of these displays? If the form is linear, approximate the r -value to express Direction and Strength. **Reminder:** An r -value close to -1 is a strong negative relationship, an r -value close to 0 is weak, and an r -value close to +1 is a strong positive!



Correlation Does Not Imply Causation!

Here are some possible correlations and the nonsense headlines a confused journalist might report as a result. In reality, the correlations have absolutely no causal relationship; they come about because both of them are related to another variable that's lurking in the background.

Can you think of another variable for each situation that might be the actual cause of the correlation and explain why the headlines the paper ran based on the correlations are nonsense?

Correlation: For a certain psychology test, the amount of time a student studied was negatively correlated with their score!

Headline: "Students who study less do better!"

Correlation: Weekly data gathered in a city throughout the year showed a positive correlation between ice cream consumption and drowning deaths.

Headline: "Drowning Deaths are Good for Ice Cream Sales!"

Correlation: A negative correlation was found between how much time students talked on the phone and how much they weighed.

Headline: "Students Lose Weight by Talking on the Phone!"

Correlations in the Animals Dataset

1) In the Interactions Area, create a scatter plot for the Animals Dataset, using "pounds" as the xs and "weeks" as the ys.

- **Form:** Does the point cloud appear linear or non-linear? _____
- **Direction:** If it's linear, does it appear to go up or down as you move from left to right? _____
- **Strength:** Is the point cloud tightly packed, or loosely dispersed? _____
- Would you predict that the r -value is positive or negative? _____
- Will it be closer to zero, closer to ± 1 , or in between? _____
- What r -value, does Pyret compute when you type `r-value(animals-table, "pounds", "weeks")`? _____
- Does this match your predictions? _____

2) In the Interactions Area, create a scatter plot for the Animals Dataset, using "age" as the xs and "weeks" as the ys.

- **Form:** Does the point cloud appear linear or non-linear? _____
- **Direction:** If it's linear, does it appear to go up or down as you move from left to right? _____
- **Strength:** Is the point cloud tightly packed, or loosely dispersed? _____
- Would you predict that the r -value is positive or negative? _____
- Will it be closer to zero, closer to ± 1 , or in between? _____
- What r -value does Pyret compute? _____
- Does this match your prediction? _____

3) Is this correlation **stronger** or **weaker** than the correlation for "pounds" ? _____

- What does that *mean* ? _____
- _____
- _____
- _____

Correlations in My Dataset

1) There may be a correlation between _____ and _____.

column

column

I think it is a _____, _____ correlation,

strong/weak

positive/negative

because _____

It might be stronger if I looked at _____

a sample or extension of my data

2) There may be a correlation between _____ and _____.

column

column

I think it is a _____, _____ correlation,

strong/weak

positive/negative

because _____

It might be stronger if I looked at _____

a sample or extension of my data

3) There may be a correlation between _____ and _____.

column

column

I think it is a _____, _____ correlation,

strong/weak

positive/negative

because _____

It might be stronger if I looked at _____

a sample or extension of my data

4) There may be a correlation between _____ and _____.

column

column

I think it is a _____, _____ correlation,

strong/weak

positive/negative

because _____

It might be stronger if I looked at _____

a sample or extension of my data

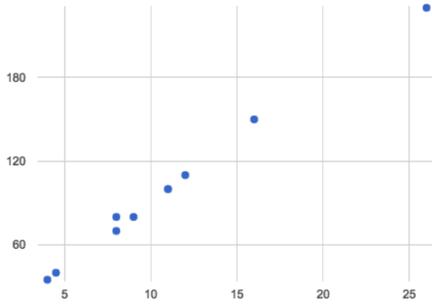
Linear Regression

- **We compute linear relationships to predict the future!** Well...sort of. Given a dataset, like ages of animals v. how long before they're adopted, we try to compute the relationship between `age` and `weeks` so that we can *predict* how long a new animal might stay, based on their age.
- When we compute linear relationships, we're talking about **straight-line patterns** that appear on a scatter plot.
- A scatter plot has an x- and y-axis, which get special names when looking for relationships. The y-axis is called the response variable, and the x-axis is called the explanatory variable. In our example, we are trying to figure out how much of the `weeks` variable is *explained by* the `age` variable.
- Linear Regression* is a way of computing the **line of best fit**, which tries to draw a line as close as possible to all the points. (Want details? It minimizes the *sum of the squares* of the vertical distances from the points to the line. There's a reason we use computers to do this!)
- **Slope** is how much we predict the response variable will increase or decrease for each unit that the explanatory variable increases. In our example, a slope of 0.5 would mean "we predict that each additional year of age means an extra half-week in the shelter". (What would a slope of 3 mean?)
- **Sample size matters!** The number of data values is also relevant. We'd be more convinced of a positive relationship in general between cat age and time to adoption if a correlation of +0.57 were based on 50 cats instead of 5.

Drawing Predictors

For each of the scatter plots below, draw a **predictor line** that seems like the best fit. Describe the correlation in terms of Direction and Strength, then estimate the r -value as being close to -1, -0.5, 0, +0.5, or +1.

A

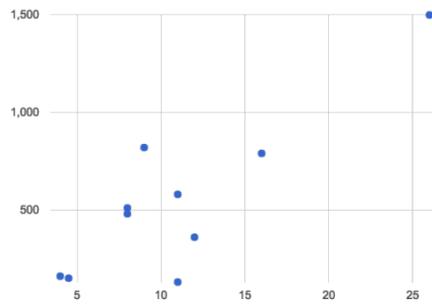


Direction: Positive Negative None

Strength: Strong Weak

r: _____

B

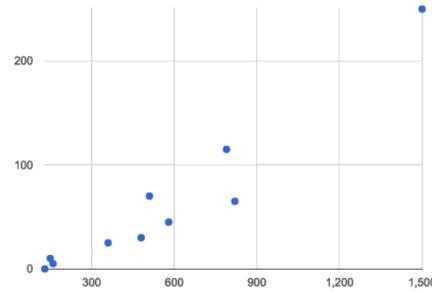


Direction: Positive Negative None

Strength: Strong Weak

r: _____

C

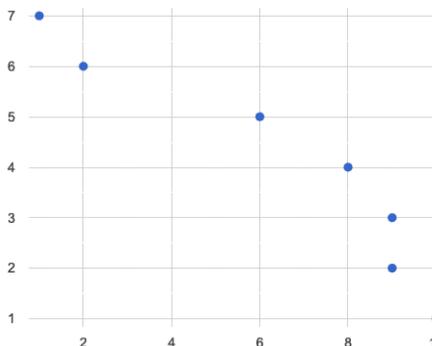


Direction: Positive Negative None

Strength: Strong Weak

r: _____

D

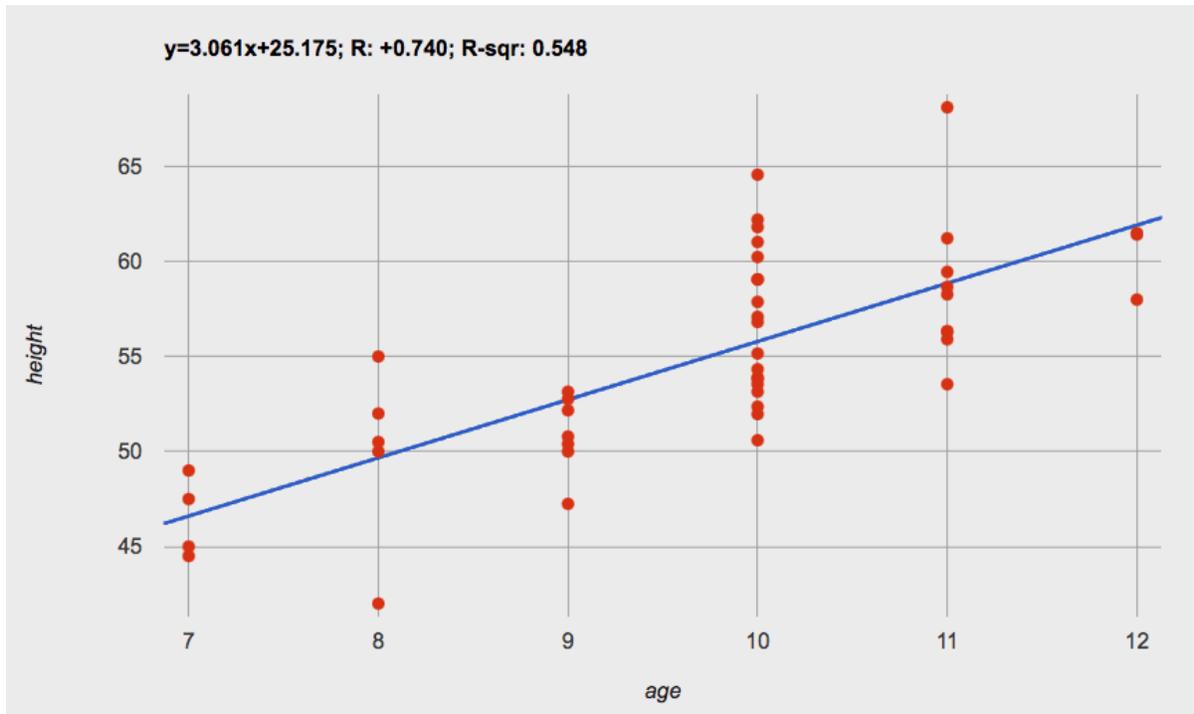


Direction: Positive Negative None

Strength: Strong Weak

r: _____

Which Questions Make Sense to Ask?



1) About how many inches are kids in this dataset expected to grow per year? _____

2) At that rate, if a child were 45" tall at age eight, how tall would you expect them to be at age twelve? _____

3) At that rate, if a ten-year-old were 55" tall at age 11, how tall would you expect them to have been at age 9? _____

4) Using the equation, how tall would you expect a seven-year-old child to be? _____

5) How many of the seven-year-olds in this sample are actually that height? _____

6) Using the equation, determine the expected height of someone who is...

7.5 years old	13 years old	6 years old	newborn	90 years old

7) For which ages is this predictor function likely to be the **most** accurate? Why?

8) For which ages is this predictor function likely to be the **least** accurate? Why?

Interpreting Regression Lines & r-Values

Each description on the left is written about the linear regression findings on the right. Fill in the blanks using the information in the line of best fit and the r-value.

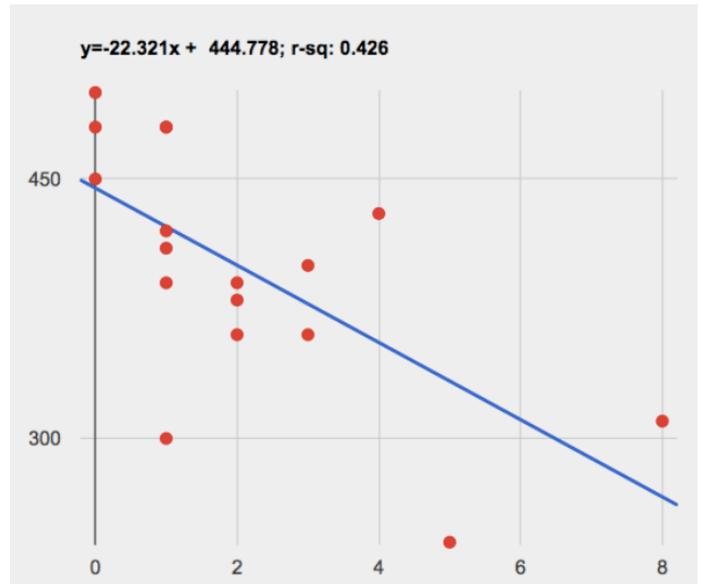
<p>1 For every additional Marvel Universe movie released each year, the average person is predicted to consume _____ pounds of sugar! This correlation is _____.</p> <p style="text-align: center;">[amount] [more / fewer]</p> <p style="text-align: center;">[strong, moderate, weak, practically non-existent]</p>	$f(x) = -3.19x + 12$ $r = -0.05$
<p>2 Shoe size and height are _____, _____ correlated. If person A is one size bigger than person B, we predict that they will be roughly _____ inches taller than person B as well.</p> <p style="text-align: center;">[strongly, moderately, weakly, not] [positively / negatively]</p> <p style="text-align: center;">[amount]</p>	$f(x) = y = 1.65x + 52$ $r = 0.89$
<p>3 There is _____ relationship found between the number of Uber drivers in a city and the number of babies born each year.</p> <p style="text-align: center;">[a strong, a moderate, no]</p>	$f(x) = 0.012x + 7.8$ $r = 0.01$
<p>4 The correlation between weeks-of-school-missed and SAT score is _____ and _____. For every week a student misses, we predict a _____ point _____ in their SAT score.</p> <p style="text-align: center;">[strong, moderate, weak, practically non-existent] [positive / negative]</p> <p style="text-align: center;">[amount] [gain / drop]</p>	$f(x) = -15.3x + 1150$ $r = -0.65$
<p>5 There is a _____, _____ correlation between the number of streaming video services someone has, and how much they weigh. For each service, we expect them to be roughly _____ pounds heavier.</p> <p style="text-align: center;">[strong, moderate, weak, practically non-existent] [positive / negative]</p> <p style="text-align: center;">[amount]</p>	$f(x) = 1.6x + 160$ $r = 0.12$

Describing Relationships

A small sample of people were surveyed about their coffee drinking and sleeping habits. Does drinking coffee impact one's amount of sleep?

NOTE: this data is made up for instructional purposes!

Daily Cups of Coffee	Sleep (minutes)
3	400
0	480
8	310
1	300
1	390
2	360
1	410
0	500
2	390
1	480
3	360
4	430
0	450
5	240
1	420
2	380
1	480



1) Describe the relationship between coffee intake and minutes of sleep shown in the data above.

2) Why is the y-axis of the display above misleading?

Data Cycle: Regression Analysis (Example)

<p>Ask Questions</p> 	<p><i>How big of a factor is age in determining adoption time for cats?</i></p> <p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p><i>cats at the shelter</i></p> <p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p><i>name, age, and weeks</i></p> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>
<p>Analyze Data</p> 	<p><i>fun is-cat(r: r["species"] == "cat" end</i></p> <p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p><i>lr-plot(animals-table.filter(is-cat), "name", "age", "weeks")</i></p> <p>What code will make the table or display you want?</p>
<p>Interpret Data</p> 	<p><i>The slope is positive (0.23) and R=0.566, so there is a moderate, positive relationship</i></p> <p>What did you find out? What can you infer?</p> <hr/> <p><i>Is the relationship different for other animals?</i></p> <p>What new question(s) does this raise?</p> <hr/>

This was a statistical question.

Write your Data Story below:

I performed a linear regression on a sample of _____ cats at the shelter _____ and found
[dataset or subset]

_____ moderate, positive _____ correlation between
a weak / strong / moderate (R=...), positive / negative

_____ age _____ and
[x-axis]

_____ time to adoption _____ . I would predict that a 1 _____ year _____ increase in
[y-axis] [x-axis units]

_____ age _____ is associated with a _____ .23 week _____ increase _____ in
[x-axis] [slope, y-units] increase / decrease

_____ time to adoption _____ .
[y-axis]

Data Cycle: Regression Analysis

<p>Ask Questions</p> 	<p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>
<p>Analyze Data</p> 	<p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p>What code will make the table or display you want?</p>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/>

Write your Data Story below:

I performed a linear regression on a sample of _____ and found
[dataset or subset]

_____ correlation between
a weak / strong / moderate (R=...), positive / negative

_____ and
[x-axis]

_____. I would predict that a 1 _____ increase in
[y-axis] [x-axis units]

_____ is associated with a _____ in
[x-axis] [slope, y-units] increase / decrease

_____.

[y-axis]

Data Cycle: Regression Analysis 2

<p>Ask Questions</p> 	<p>What question do you have?</p> <hr/> <p>What type of question is this? (circle one) Lookup Arithmetic Statistical</p>
<p>Consider Data</p> 	<p>What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)</p> <hr/> <p>What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)</p>
<p>Analyze Data</p> 	<p>If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)</p> <hr/> <p>If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)</p> <hr/> <p>What code will make the table or display you want?</p>
<p>Interpret Data</p> 	<p>What did you find out? What can you infer?</p> <hr/> <p>What new question(s) does this raise?</p> <hr/>

Write your Data Story below:

I performed a linear regression on a sample of _____ and found
dataset or subset

_____ correlation between
a weak/strong/moderate (R=...), positive/negative

_____ and
[x-axis]

_____. I would predict that a 1 _____ increase in
[y-axis] [x-axis units]

_____ is associated with a _____ in
[x-axis] [slope, y-units] [increase/decrease]

_____.

[y-axis]

Katyanna Quach Wed 22 May 2019 // 23:50 UTC

ANALYSIS AI experts, lawyers, and law enforcement urged US Congress to regulate the use of facial recognition technology during a hearing held by the House Committee on Oversight and Reform on Wednesday, May 22, 2019.

The technical issues and social impacts of using AI software to analyse images or videos are well known. There have been repeated reports of how inaccuracies lead to people being misidentified in research and in real life. San Francisco just passed an ordinance banning the local government using facial recognition technology.

In some cases, like the experiment conducted by the American Civil Liberties Union's (ACLU), a nonprofit based in New York, that showed Amazon Rekognition incorrectly matched members of the US Congress to criminal mugshots, the effects have been minimal. It's simply absurd for elected politicians to be wanted criminals. But what happens when the technology is turned on civilians who have less power?

At a hearing of the House Committee on Oversight and Reform on Wednesday, Joy Buolamwini, founder of Algorithmic Justice League, an activist collective focused on highlighting the shortcomings of facial recognition, found that commercial computer models struggled most when it came to recognizing women with darker skin. IBM's system was incorrect for 34.7 per cent of the time when it came to identifying black women, she said...

The problem boiled down to biased training datasets, Buolamwini told the House committee. AI systems perform worse on data that they haven't seen before. So, if most datasets mainly represent white men then it's not surprising that they find it difficult when faced with an image of women of colour.

When it comes to databases of mugshots, however, the reverse is true. Black people are overrepresented in mugshot databases, explained Clare Garvie, Senior Associate at Georgetown University Law Center's Center on Privacy & Technology. If law enforcement are using these flawed models to target the group of people that it struggles to identify most then it will undoubtedly lead to police stopping and searching the wrong people. "It's a violation of the first and fourth amendment," Garvie said during the hearing.

Law enforcement and lack of transparency

Cedric Alexander, the former president of the National Organization of Black Law Enforcement Executives who was also a witness at the hearing, estimated that at least a quarter of law enforcement agencies across the US use facial recognition to some degree.

Police from Washington County and Orlando are an example of some bureaus that are using Rekognition. Michael Punke, Amazon's VP of Global Public Policy, said at the time it has "not received a single report of misuse by law enforcement." It's difficult to verify that claim, however, considering that the police haven't been transparent about how it's used.

It's all done in secrecy, according to testimony. Elijah Cummings, the chair of the Oversight Committee, said that 18 states had shared data like passport photos or driver licenses with the FBI without explicit consent. When the witnesses were pressed with questions on what kind of information law agencies share with one another, nobody knew.

Neema Guliani, senior legislative counsel for the ACLU, took a tough stance and called for a moratorium on the technology. She urged the committee to "take steps to halt the use of face recognition for law enforcement and immigration enforcement purposes until Congress passes a law dictating what, if any, uses are permissible and ensures that individuals' rights can be protected." Unregulated use of the technology could also potentially lead to an "Orwellian surveillance state," where citizens are constantly tracked Guliani said. In the opening statement, Cummings said there are about 50 million surveillance cameras in the US, and that half of all American adults are probably part of facial recognition databases and they don't even know it.

Andrew Ferguson, professor of law at the University of the District of Columbia, agreed that the Congress needed to act now to prohibit facial recognition until Congress establishes clear rules. "Unregulated facial recognition should not be allowed to continue unregulated. It is too chilling, too powerful. The fourth amendment won't save us. The Supreme Court is trying to make amendments but it's not fast enough. Only legislation can react in real time to real time threats," he warned.

Alexander was more cautious about a blanket ban on the technology, however. He believed that there were still ways that law enforcement could positively use facial recognition. "There is a place for the technology, but the police need to be trained properly. They can't just be passed the technology by software companies." Effective policing is about building relationships in the local community, and it can't afford the effects of misidentifying people. How can we utilise the technology, whilst developing some standards?, he asked.

Benchmark tests simply aren't good enough

The National Institute of Standards and Technology (NIST), a laboratory part of the US Department of Commerce, is currently conducting official benchmark tests for commercial facial recognition systems. But they need to be better, Buolamwini said. She brought up the issue of what she called "pale male datasets". "The gold standard benchmark dataset is biased and can lead to a false understanding of progress," she said.

Even if there was a facial recognition system with near-perfect accuracy in the testing phase, it doesn't solve the problem that most data used by law enforcement is often grainy and low resolution. A recent report by Georgetown University found that in some cases police were even trying to match people by composite artist sketches.

"Faces maybe the final frontier of privacy," Buolamwini said.

The hearing took place at the same time as Amazon shareholders tried to stop Rekognition being sold to law enforcement. The proposal was defeated, but the vote tallies were not immediately disclosed. © **The Register.**

Can Software be Biased?

This page is designed as a reflection on either [this article, summarizing US Congress Testimony on Artificial Intelligence](#) or this video [The Coded Gaze: Bias in Artificial Intelligence](#).

1) Describe three concerns experts and activists have raised about Artificial Intelligence.

2) What are some solutions that would address these concerns?

3) How would you test whether or not a facial recognition system was equally accurate for everyone?

Threats to Validity

Threats to Validity can undermine a conclusion, even if the analysis was done correctly.

Some examples of threats are:

- **Selection bias** - identifying the favorite food of the rabbits won't tell us anything reliable about what all the animals eat.
- **Study bias** - If someone is supposed to assess how much cat food is eaten each day on average, but they only measure how much cat food is put in the bowls (instead of how much is actually consumed), they'll end up with an over-estimate.
- **Poor choice of summary** - Suppose a different shelter that had 10 animals recorded adoption times (in weeks) as 1, 1, 1, 7, 7, 8, 8, 9, 9, 10. Using the mode (1) to report what's typical would make it seem like the animals were adopted much quicker than they really were, since 7 out of 10 animals took at least 7 weeks to be adopted.
- **Confounding variables** - Some shelter workers might prefer cats, and steer people towards cats as a result. This would make it appear that "cats are more popular with people", when the real variable dominating the sample is what *workers at the shelter* prefer.

Identifying Threats to Validity

Some volunteers from the animal shelter surveyed a group of pet owners at a local dog park. They found that almost all of the owners were there with their dogs. From this survey, they concluded that dogs are the most popular pet in the state.

What are some possible threats to the validity of this conclusion?

The animal shelter noticed a large increase in pet adoptions between Christmas and Valentine's Day. They conclude that at the current rate, there will be a huge demand for pets this spring.

What are some possible threats to the validity of this conclusion?

Identifying Threats to Validity (2)

The animal shelter wanted to find out what kind of food to buy for their animals. They took a random sample of two animals and the food they eat, and they found that spider and rabbit food was by far the most popular cuisine!

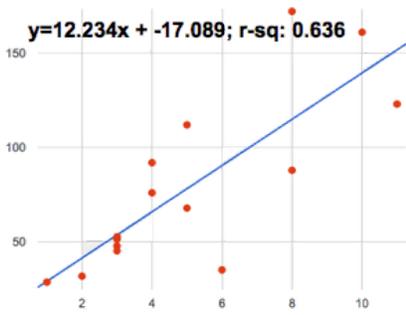
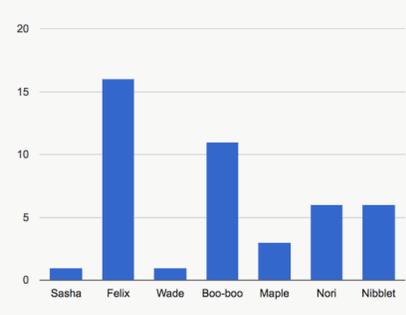
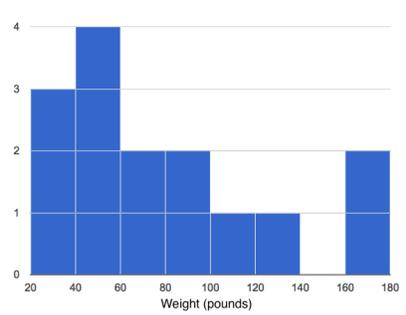
Explain why sampling just two animals can result in unreliable conclusions about what kind of food is needed.

A volunteer opens the shelter in the morning and walks all the dogs. At mid-day, another volunteer feeds all the dogs and walks them again. In the evening, a third volunteer walks the dogs a final time and closes the shelter. The volunteers report that the dogs are much friendlier and more active at mid-day, so the shelter staff assume the second volunteer must be better with animals than the others.

What are some possible threats to the validity of this conclusion?

Fake News

There are six separate, *unrelated* claims below, and ALL OF THEM ARE WRONG! Your job is to figure out why by looking at the data.

	Data	Claim	What's Wrong
1	The average player on a basketball team is 6'1".	"Most of the players are taller than 6'."	
2	Linear regression found a positive correlation ($r=0.42$) between people's height and salary.	"Taller people are more qualified for their jobs."	
3		"According to the predictor function indicated here, the value on the x-axis will predict the value on the y-axis 63.6% of the time."	
4		"According to this bar chart, Felix makes up a little more than 15% of the total ages of all the animals in the dataset."	
5		"According to this histogram, most animals weigh between 40 and 60 pounds."	
6	Linear regression found a negative correlation ($r= -0.91$) between the number of hairs on a person's head and their likelihood of owning a wig.	"Owning wigs causes people to go bald."	

Lies, Darned Lies, and Statistics

- 1) Using real data and displays from your dataset, come up with a misleading claim.
- 2) Trade papers with someone and figure out why their claims are wrong!

Data	Claim	Why it's wrong
3		
4		
5		
6		

Data Cycle

Ask Questions



What question do you have?

What type of question is this? (circle one) Lookup Arithmetic Statistical

Consider Data



What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)

What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)

Analyze Data



If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)

If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)

What code will make the table or display you want?

Interpret Data



What did you find out? What can you infer?

What new question(s) does this raise?

Ask Questions



What question do you have?

What type of question is this? (circle one) Lookup Arithmetic Statistical

Consider Data



What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)

What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)

Analyze Data



If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)

If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)

What code will make the table or display you want?

Interpret Data



What did you find out? What can you infer?

What new question(s) does this raise?

Data Cycle

Ask Questions



What question do you have?

What type of question is this? (circle one) Lookup Arithmetic Statistical

Consider Data



What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)

What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)

Analyze Data



If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)

If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)

What code will make the table or display you want?

Interpret Data



What did you find out? What can you infer?

What new question(s) does this raise?

Ask Questions



What question do you have?

What type of question is this? (circle one) Lookup Arithmetic Statistical

Consider Data



What Rows should we investigate? (All the rows, just the cats, fixed dogs, etc.)

What Column(s) do we need? (age, weight-in-kilograms, weeks, etc.)

Analyze Data



If you only need some rows, define your filter function here (Need help? Use the Design Recipe!)

If you need to make a new column, define your builder function here (Need help? Use the Design Recipe!)

What code will make the table or display you want?

Interpret Data



What did you find out? What can you infer?

What new question(s) does this raise?

Design Recipe

Directions:

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

what the function does with those variable(s)

end

Directions:

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

what the function does with those variable(s)

end

Design Recipe

Directions:

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

what the function does with those variable(s)

end

Directions:

Contract and Purpose Statement

Every contract has three parts...

_____ :: _____ -> _____
function name Domain Range

what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) is _____
function name input(s) what the function produces

_____ (_____) is _____
function name input(s) what the function produces

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name variable(s)

what the function does with those variable(s)

end

Contracts

Contracts tell us how to use a function. For example: `num-min :: (a :: Number, b :: Number) -> Number` tells us that the name of the function is `num-min`, it takes two inputs (both Numbers), and it evaluates to a Number. From the contract, we know `num-min(4, 6)` will evaluate to a Number. Use the blank line under each contract for notes or sample code for that function!

Name	Domain	Range
<code># num-sqr</code>	<code>:: Number</code>	<code>-> Number</code>
<code>num-sqr (9)</code>		
<code># num-sqrt</code>	<code>:: Number</code>	<code>-> Number</code>
<code>num-sqrt (25)</code>		
<code># triangle</code>	<code>:: Number, String, String</code>	<code>-> Image</code>
<code>triangle (80, "solid", "darkgreen")</code>		
<code># circle</code>	<code>::</code>	<code>-> Image</code>
<code># star</code>	<code>::</code>	<code>-></code>
<code># square</code>	<code>::</code>	<code>-></code>
<code># rectangle</code>	<code>::</code>	<code>-></code>
<code># text</code>	<code>::</code>	<code>-></code>
<code># ellipse</code>	<code>::</code>	<code>-></code>

Contracts

Contracts tell us how to use a function. For example: `num-min :: (a :: Number, b :: Number) -> Number` tells us that the name of the function is `num-min`, it takes two inputs (both Numbers), and it evaluates to a Number. From the contract, we know `num-min(4, 6)` will evaluate to a Number. **Use the blank line under each contract for notes or sample code for that function!**

Name	Domain	Range
<code># regular-polygon</code>	<code>::</code>	<code>-></code>
<code># rhombus</code>	<code>::</code>	<code>-></code>
<code># right-triangle</code>	<code>::</code>	<code>-></code>
<code># isosceles-triangle</code>	<code>::</code>	<code>-></code>
<code># radial-star</code>	<code>::</code>	<code>-></code>
<code># star-polygon</code>	<code>::</code>	<code>-></code>
<code># overlay</code>	<code>:: Image, Image</code>	<code>-> Image</code>
<code>overlay(star(30, "solid", "gold"), circle(30, "solid", "blue"))</code>		
<code># beside</code>	<code>:: Image, Image</code>	<code>-> Image</code>
<code>beside(star(50, "solid", "orange"), circle(50, "solid", "green"))</code>		
<code># above</code>	<code>:: Image, Image</code>	<code>-> Image</code>
<code>above(triangle(30, "solid", "red"), square(30, "solid", "blue"))</code>		

Contracts

Contracts tell us how to use a function. For example: `num-min :: (a :: Number, b :: Number) -> Number` tells us that the name of the function is `num-min`, it takes two inputs (both Numbers), and it evaluates to a Number. From the contract, we know `num-min(4, 6)` will evaluate to a Number. Use the blank line under each contract for notes or sample code for that function!

Name	Domain	Range
<code># put-image</code>	<code>:: Image, Number, Number, Image</code>	<code>-> Image</code>
<code>put-image(star(30, "solid", "red"), 50, 150, rectangle(300, 200, "outline", "black"))</code>		
<code># rotate</code>	<code>:: Number, Image</code>	<code>-> Image</code>
<code>rotate(35, rectangle(30, 80, "solid", "purple"))</code>		
<code># scale</code>	<code>:: Number, Image</code>	<code>-> Image</code>
<code>scale(0.8, triangle(30, "solid", "red"))</code>		
<code># string-repeat</code>	<code>:: String, Number</code>	<code>-> String</code>
<code>string-repeat("cheetah", 5)</code>		
<code># string-contains</code>	<code>:: String, String</code>	<code>-> Boolean</code>
<code>string-contains("rockstar", "star")</code>		
<code># num-min</code>	<code>:: Number, Number</code>	<code>-> Number</code>
<code>num-min(80, 20)</code>		
<code># num-max</code>	<code>:: Number, Number</code>	<code>-> Number</code>
<code>num-max(80, 20)</code>		
<code># count</code>	<code>:: Table, String</code>	<code>-> Table</code>
<code>count(animals-table, "species")</code>		
<code># mean</code>	<code>:: Table, String</code>	<code>-> Number</code>
<code>mean(animals-table, "age")</code>		

Contracts

Contracts tell us how to use a function. For example: `num-min :: (a :: Number, b :: Number) -> Number` tells us that the name of the function is `num-min`, it takes two inputs (both Numbers), and it evaluates to a Number. From the contract, we know `num-min(4, 6)` will evaluate to a Number. **Use the blank line under each contract for notes or sample code for that function!**

Name	Domain	Range
<code># median</code>	<code>:: Table, String</code>	<code>-> Number</code>
<code>median(animals-table, "age")</code>		
<code># modes</code>	<code>:: Table, String</code>	<code>-> List<Number></code>
<code>modes(animals-table, "age")</code>		
<code># bar-chart</code>	<code>:: Table, String</code>	<code>-> Image</code>
<code>bar-chart(animals-table, "legs")</code>		
<code># pie-chart</code>	<code>:: Table, String</code>	<code>-> Image</code>
<code>pie-chart(animals-table, "species")</code>		
<code># histogram</code>	<code>:: (t :: Table, column :: String, bin-width :: Number)</code>	<code>-> Image</code>
<code>histogram(animals-table, "age", 2)</code>		
<code># box-plot</code>	<code>:: Table, String</code>	<code>-> Image</code>
<code>box-plot(animals-table, "age")</code>		
<code># modified-box-plot</code>	<code>:: Table, String</code>	<code>-> Image</code>
<code>modified-box-plot(animals-table, "age")</code>		
<code># scatter-plot</code>	<code>:: (t :: Table, labels :: String, xs :: String, ys :: String)</code>	<code>-> Image</code>
<code>scatter-plot(animals-table, "species", "pounds", "weeks")</code>		
<code># image-scatter-plot</code>	<code>:: (t :: Table, xs :: String, ys :: String, f :: (Row -> Image))</code>	<code>-> Image</code>
<code>image-scatter-plot(animals-table, "pounds", "weeks", animal-img)</code>		

Contracts

Contracts tell us how to use a function. For example: `num-min :: (a :: Number, b :: Number) -> Number` tells us that the name of the function is `num-min`, it takes two inputs (both Numbers), and it evaluates to a Number. From the contract, we know `num-min(4, 6)` will evaluate to a Number. **Use the blank line under each contract for notes or sample code for that function!**

Name	Domain	Range
<code># r-value</code>	<code>(t :: Table, xs :: String, ys :: String)</code>	<code>-> Number</code>
<code>r-value (animals-table, "pounds", "weeks")</code>		
<code># lr-plot</code>	<code>(t :: Table, labels :: String, xs :: String, ys :: String)</code>	<code>-> Image</code>
<code>lr-plot (animals-table, "species", "pounds", "weeks")</code>		
<code># random-rows</code>	<code>(t :: Table, num-rows :: Number)</code>	<code>-> Table</code>
<code>random-rows (animals-table, 5)</code>		
<code># <Table>.row-n</code>	<code>Number</code>	<code>-> Row</code>
<code>animals-table.row-n(5)</code>		
<code># <Table>.order-by</code>	<code>(col :: String, increasing :: Boolean)</code>	<code>-> Table</code>
<code>animals-table.order-by("species", true)</code>		
<code># <Table>.filter</code>	<code>(test :: (Row -> Boolean))</code>	<code>-> Table</code>
<code>animals-table.filter(is-cat)</code>		
<code># <Table>.build-column</code>	<code>(col :: String, builder :: (Row -> Any))</code>	<code>-> Table</code>
<code>animals-table.build-column("sticker", label)</code>		
<code># bar-chart-summarized</code>	<code>(t :: Table, labels :: String, values :: String)</code>	<code>-> Image</code>
<code>bar-chart-summarized(animals-table, "species", "pounds")</code>		
<code># pie-chart-summarized</code>	<code>(t :: Table, labels :: String, values :: String)</code>	<code>-> Image</code>
<code>pie-chart-summarized(animals-table, "age", "pounds")</code>		



These materials were developed partly through support of the National Science Foundation, (awards 1042210, 1535276, 1648684, and 1738598), and are licensed under a Creative Commons 4.0 Unported License. Based on a work at www.BootstrapWorld.org. Permissions beyond the scope of this license may be available by contacting contact@BootstrapWorld.org.