





Workbook v1.0

Brought to you by the Bootstrap team:

- Emmanuel Schanzer
- Kathi Fisler
- Shriram Krishnamurthi
- Sam Dooman
- Ed Campos

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 schanzer@BootstrapWorld.org.

Unit 1

- 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 of 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 data usually numeric that measures *quantity*, such as a person's height, a score on test, a measure of distance, etc. A list of quantitative data can be ordered from smallest to largest.
- Categorical data is data that specifies categories, such as eye color, country of origin, etc. A list of categorical data has no notion of "smallest" or "largest", and cannot be ordered.
- **Programming languages** involves different *datatypes*, such as Numbers, Strings, Booleans and Images.
- Operators (like +, -, *, <, etc.) are written between values. For example: 4 + 2
- **Functions** (like triangle, star, string-repeat, etc.) are written first, followed by a list of **arguments** in parentheses. For example: star(50, "solid", "red")

Numbers and Strings

Make sure you've loaded the Unit 1 Starter File, and clicked "Run".

- 1. Try typing 42 into the Interactions Area and hitting "Enter". What happens?
- 2. Try typing in other Numbers. What happens if you try a decimal like 0.5? A fraction like 1/3? Try really big Numbers, and really small ones.
- 3. String values are always in quotes. Try typing your name in quotes, and see what happens when you hit "Enter".
- 4. Try typing your name without the closing quote. What happens? Now try typing it without any quotes.
- 5. Is 42 the same as "42"? Why or why not? Write your answer below:
- 6. Just like in math, Pyret has operators like + and -. Try typing in 4 + 2, and then 4+2 (without the spaces). What can you conclude from this? Write your answer below:
- 7. Try typing in 4+2+6, 4+2*6, and 4+(2*6). What can you conclude from this? Write your answer below:
- 8. Try typing in 4 + "cat", and then "dog" + "cat". What can you conclude from this? Write your answer below:

Booleans

Boolean expressions are yes-or-no questions, and you probably already know some Boolean operators from math class, which compare Numbers. What will each of the shaded expressions below evaluate to? Type them into Pyret if you're not sure.

3 <= 4	 "a" > "b"	
3 == 2	 "a" <> "b"	
2 <> 4	 "a" == "b"	
3 <> 3	 "a" <> "a"	

Boolean Operators

Pyret also has operators that work on *Booleans*. For each expression below, write down your guess about what it will evaluate to. Then type them in and see if you were right!

- 2. How many different String values are there in Pyret?
- 3. How many different Boolean values are there in Pyret? _____

Unit 2

- Programming languages let us **define our own function**.
- We use the **Design Recipe** to help us define functions without making mistakes.
- The first step is to write a **Contract** and **Purpose Statement** for the function, which specify the Name, Domain and Range of the function and give a summary of what it does.
- The second step is to **write at least two examples**, which show how the function should work for specific inputs. These examples help us see patterns, and we express those patterns by **circling and labeling** what changes.
- The final step is to **define the function**, which generalizes our examples.

The Animals Dataset

1.	My dataset is	Animals from a pet store
_		

2.	Some	of the	columns	in	my	dataset	are	:
----	------	--------	---------	----	----	---------	-----	---

Name (capitalization matters!)	Datatype	Quantitative/Categorical

. 1	Three questions I have about the an	nimals dataset:	
	2.		
	3.		

The Design Recipe

Define a function called is-fixed, which tells us whether or not an animal is fixed is-fixed (animal :: Row) Boolean domain range # Consumes an animal, and produces the value in the fixed column examples: is-fixed (__sasha__) is _____ true (____) is ___ end fun end Define a function called gender, which consumes a Row of the animals table tells us the gender of that animal domain name range examples: () **is**

end

fun

end

) is

Define a function called is-cat,	which consumes of	Row of the	animals t	able d	and
produces true if it's a cat.					

	is-cat	:	(animal :: Row)	\rightarrow	Boolean
	name		domain		range
# Cor	nsumes an anim	al, and retur	n true if the species	is "cat"	
" —			<u>`</u>		
exam	ples:				
	is-cat	(sash	a		
	15-641	(<u>a</u>) is		
		1) is		
end					
Ciiu					
fun		() :		
end					
			ng, which consumes		inimais table and
produc	ces true ii ii s	an animai ir	at is less than two y	ears ola.	
		:		\rightarrow	
	name		domain		range
#					
 exam	ples:				
	F				
		() is		
		() is		
end		`	/		
fun		() :		
end					

Defin	e a function cal	led namet	ag, prints out each ani	mal's name in b	oig red letters.
	nametag	<u>:</u>	(animal :: Row)	>	Image
	name		domain		range
# Co	nsumes an anima	al, and prod	duces an image of their	name in big, red	l letters
exar	mples:			-	
	nametag	(sha) is		
•		() is		
end					
fun		() :		
end					
	e a function cal uces true if it's a		tten, which consumes	a Row of the c	ınimals table and
				\rightarrow	
#	name	•	domain	/	range
exar	mples:				
		() is		
		() is		
end					
fun		() :		
end					

My Dataset

1. My dataset is				
2. Four of my column (choose four columns, Qualitative or Categor	and for each one fil		tatype, and whethe	r it contains
Name				
Datatype				
Quantitative/Categorica	1			
3. Three questions I ho	ave about my do	ataset:		
2.				
3.				

Unit 3

- Methods are special functions that are attached to pieces of data. We use them to manipulate Tables.
- They are different from functions in several ways:
 - 1. Their names can't be used alone: they can only be used as part of data, separated by a dot. (For example, animals.order-by)
 - 2. Their contracts are different: they include the type of the data as part of their names. (eg, .order-by :: (column :: String) → Table)
 - 3. They have a "secret" argument, which is the data they are attached to
- We will use three **Table Methods** to manipulate our datasets:
 - 1. <Table>.order-by order the rows of a table based on a column
 - 2. <Table>.filter create a **subset** of the data, with only certain rows
 - 3. <Table>.build-column use the columns of a table to compute a new one
- We use Table Plans to help us use these methods correctly, without making mistakes

Reviewing Functions

1.	How many functions are defined in this file?	
2.	What is the name of the last function?	
3.	What is the Domain of the last function?	
4.	What is the Range of the last function?	
5.	What is the Range of the last function?	
6.	What is the variable name that the last function uses?	
7.	Which function will tell us if an animal is a kitten?	
8.	Which function will print out " <name> the <species>"?</species></name>	
9.	Which function will tell us if an animal is a dog older than 10?	
10	.Which function will tell us if an animal has been fixed?	
11	. Which function will draw a nametag for an animal?	

12. One of the examples for the last function is broken. Fix this example in the Definitions Area.

Plans for the Animals Dataset

What are two ways you might want to order the animals dataset?
1)
2)
What are two subsets into which you might filter the animals dataset?
1)
2)
What are two new columns you might want to build from the animals dataset?
1)
2)

Methods

Methods are a lot like functions, but they differ in three important ways:

- They can only be called as **part of a value**, using the **dot-accessor**. For example: **animals.**row-n(2)
- Their Contracts are different, because they contain a **Type** as part of their name. For example: <**Table>.**row-n :: (index :: Number) -> Row
- They have a "secret argument", which is the value they are attached to. In the examples above, the row-n method consumes only a Number as part of its Domain, but it also consumes a Table.

Here is the Contract for a method, which consumes the name of a food and produces True if the person likes that food:

	<person>.likes :: (food :: String) -> Boolean</person>
1.	What Type of data is the method attached to?
2.	What is the name of this method?
3.	How many things are in its Domain?
4.	What is the name of the argument in its Domain?
5.	What is the Type of the argument in its Domain?
6.	What Type of data will this method will produce?
7.	Below are 3 expressions. Based on the contract above, circle the correct one.
	emma.likes("pizza") likes("pizza") likes(emma, pizza)

8. On the line below, write your own expression that uses this method, replacing emma and "pizza" with your own name and a food you like.

On Kitten Day, the shelter prints up a list of all the cats in their database that are less than 2 years old, and makes nametags for them. They need a function that will help them out! Define a function called get-kittens-tags, which takes in the dataset and produces the correct table.

name species age fixed Sasha cat 1 FALSE	nd a result bo	weight 6.5	I on the
name species age fixed Sasha cat 1 FALSE	fixed legs we FALSE 4	weight 6.5	adopt
ake a Start Table and a renimals—table name species age fixed Sasha cat 1 FALSE	fixed legs we FALSE 4	weight 6.5	adopt
Sasha cat 1 FALSE	fixed legs we FALSE 4	weight 6.5	adopt
namespeciesagefixedSashacat1FALSE	FALSE 4	6.5	
Sasha cat 1 FALSE	FALSE 4	6.5	
			4
Togale dog 3 TRUE	TRUE 4		
	151105 1	48	3
Buddy lizard 2 FALSE		0.3	12
Wadecat1FALSEMittenscat2TRUE		7.4	5
MILLELIS COL Z INOL	INUL 4		

The first weekend of every month, the shelter holds a "meet the dogs" picnic, to encourage families to adopt their dogs. Write a function called get-dogs-by-age, that takes their database and produces a table of all the dogs in the shelter, sorted from youngest to oldest.

Contract	and Pu	rpos	е											
get	-dogs-b	y-ag	де	:		(anima	als :: Tabi	le)	\rightarrow	•		Tai	ble	
# Consum	e a table	of ai	nimals, a	nd pr	roduce	a table	containing	only the	dogs, so	rted	by ag	e		
Example: Make a St		e anc	d a resul	t bas	sed on	that to	ıble.							
animals	-table	<u>e</u>					\rightarrow	get-d	og-by	-ag	e(an	ima	ls-ta	able)
name	species	age	fixed	legs	weight	adopt								
Snowcone	cat	2	TRUE	4	6.1	5		name	species	age	fixed	legs	weight	adopt
Wade	cat	1	FALSE	4	3.2	4		Toggle	dog	3	TRUE	4	48	3
Hercules	cat	3	FALSE	4	13.4	7		Fritz	dog	4	TRUE	4	92	6
Toggle	dog	3	TRUE	4	48	3	!		l	l				
Define th			ods Loiro	lo vo	vur holr	or fun	ctions!), th	on produ	co d ro	cult v	vith th	0.00	w table	
			•	,	·		,	en produ	ce a re:	SOII V	VIII II I	e 116	W IGDI	J.
fun					():				<u>D</u>	efir	e the	table
.bui	ild-colui	mns(,)	Are	there	more c	olumns?
)	Ar	e the	re fewe	er rows?
	der-by(Ar	e the	rows oi	rdered?
											Pro	duce	e the	result

It's important for animals to stay healthy, especially when they get older. The veterinarians at the shelter want to put some of the dogs on a diet! They need a regular report of all the older dogs, sorted from heaviest-to-lightest. Define a function old-dogs-diet, which does just that!

Examples Make a Start		nd a	result b	ased	I on the	at tab	le.		>	_				
		nd a	result b	ased	I on the	at tab	le.							
		nd a	result b	ased	I on the	at tab	le.							
		nd a	result b	ased	I on the	at tab	le.							
		nd a	result b	ased	I on the	at tab	le.							
Make a Start		nd a	result b	ased	I on the	at tab	le.							السراح
	able													
	<u>able</u>													
animals-t							\rightarrow	old-	dogs-	die	t(an	imal	s-tal	ble)
name	species	age	fixed	legs	weight	adopt								
Snowcone	cat	2	TRUE	4	6.1	5		name	species	age 3	fixed TRUE	legs 3	weight 45.4	adopt
Lucky	dog	3	TRUE	3	45.4	9		Lucky	dog cat	2	TRUE	4	6.1	5
Hercules	cat	3	FALSE	4	13.4	7		Toggle	dog	3	TRUE	4	48	3
Toggle	dog	3	TRUE	4	48	3		roggie	aog	J	IKUE	4	40	3
Snuggles ta	rantula	2	FALSE	8	0.1	1								
Define the f	unction													
Use the relev		nods	(circle	your	helper	funct	ions!),	then produ	ice a re	sult v	vith the	e new	table.	
fun					_ ()	:			6	<i>c</i> ·		, , ,
<u>† =</u>											D	etine	the t	rable
)	Are t	here n	nore col	umns?
											Are	e there	e fewer	rows?
	-by()	Are	e the r	ows ord	lered?
ui uei	- <i>Dy</i> (0	./	41-	/ 1
 end											Prod	<u>auce</u>	the re	2SUIT

The shelter is tracking birth-years for all the animals who've been fixed. They need a function that takes in their database and returns a table that contains the birth-year for each one. Define get-fixed-birth that will do this for them.

Contract	and Purpo	ose												
			:						->	>				
Example	S													
	art Table a	nd a	result b	asec	on the	at tabl	e.							
animals	-table						\rightarrow get-f	ixed-	-by-	legs	(ani	mals	<u>-tab</u>	ole)
name	species	age	fixed	legs	weight	adopt								. 1
Snowcone	cat	2	TRUE	4	6.1	5	name Snowcone	species	age 2	fixed TRUE	legs 4	weight 6.1	adopt 5	2015
Lucky	dog	3	TRUE	3	45.4	9	Lucky		3	TRUE	3	45.4	9	2013
Hercules	cat	3	FALSE	4	13.4	7	Toggle	dog	3	TRUE	4	43.4	3	2014
Toggle	dog	3	TRUE	4	48	3	Toggle	dog	3	TRUE	4	48	3	2014
Snuggles	tarantula	2	FALSE	8	0.1	1								
			1		l									
D. C	- f l'													
	e function		(circle	vour	helper	functi	ons!), then p	roduce	a re	sult wit	h the	new to	able	
			(00.0	, 00.	1101001	1011011	0113.77 111011 P	.0000	, G 10	JOII 1111		11011	<i>a</i>	
fun					1		١.							
											De	efine 1	the to	<u>able</u>
											100 +1	here moi	wa aali	
_ <i>.bui</i>	<u>ild-column</u>	<u>s(</u>)				
fil	ter()		there f		
ord	der-by(_)	Are	the row	vs orde	ered?
	<u>-</u>				<u>-</u>						Proc	duce ti	he re	sult
end														

My Dataset

What are two ways you might want to order this dataset?
1)
2)
What are two subsets into which you might filter this dataset?
1)
2)
What are two new columns you might want to build from this dataset?
_1)
2)

Unit 4

- Bar charts show the absolute quantity of each row in a dataset. The larger the
 quantity, the longer the bar. Bar charts provide a visual representation of values
 in a dataset.
- **Pie charts** show the *relative* quantity of each row in a dataset. The greater the percentage, the larger the pie slice. Pie charts provide a visual representation of proportions in a dataset.
- **Choosing a Sample Table** is important when coming up with small examples for Table Plans. A good sample table has:
 - 1. At least all the relevant columns
 - 2. Enough rows to accurately represent the dataset
 - 3. Rows that are randomly-ordered

Statements about Columns

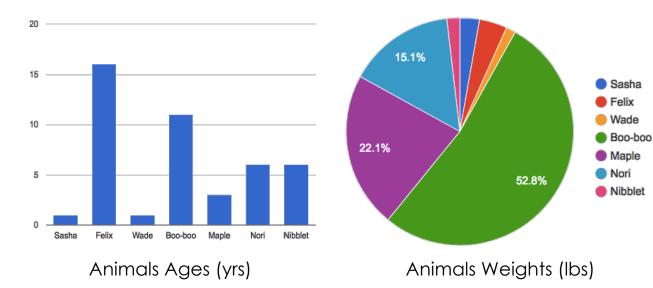
Use the Table below to help you answer the questions.

name	species	age	pounds
Sasha	cat	1	6.5
Felix	cat	16	9.2
Wade	cat	1	3.2
Boo-boo	dog	11	123
Maple	dog	3	51.6
Nori	dog	6	35.3
Nibblet	rabbit	6	4.3

Which animal(s) is/are the heaviest?
 Which animal(s) is/are the youngest?
 How much of the total weight comes from Maple?
 How much of the combined age comes from Nori?
 Would these questions be harder to answer if the table had 100 rows? If so, why?

Visualizing Quantity

In the table below, there are two observations drawn from the following charts. Add two more.



Based on a chart of	I notice that
Based on a bar chart of 7 animals' ages	Felix is by far the oldest
Based on a pie chart of 7 animals' weights	Boo-boo weighs more than the other six animals combined!
Based on a bar chart of 7 animals' ages	
Based on a pie chart of 7 animals' weights	

Dogs are generally a lot bigger heavier than cats, so the shelter wants to look at a chart of *only* the dogs to determine who needs more exercise time. Define a function pie-dog-weight, which will make a pie chart showing the relative weights of all the dogs in the shelter.

Examples Make a Start Table and a result based on that table. animals-table name			pose			
Make a Start Table and a result based on that table. animals—table name	1		:			\rightarrow
Make a Start Table and a result based on that table. animals—table name						
Make a Start Table and a result based on that table. animals—table pie-dog-weight(animals) name						
Make a Start Table and a result based on that table. animals—table pie-dog-weight(animals) name						
Make a Start Table and a result based on that table. animals—table pie-dog-weight(animals) name	Examples					
name weight Snowcone 6.1 Lucky 45.4 Hercules 13.4 Toggle 48		t Table	and a result	based on that tak	ole.	
Snowcone 6.1 Lucky 45.4 Hercules 13.4 Toggle 48	animals-	table	<u>!</u>	\rightarrow	<u>p</u> :	ie-dog-weight(animals
Lucky 45.4 Hercules 13.4 Toggle 48	name		weight			
Toggle 48	Snowcone		6.1			
Toggle 48	Lucky		45.4			
	Hercules		13.4			
Spugglos			48			
Shuggles 0.1	Snuggles		0.1			
				a vour helper func	tions) then r	produce a result with the new t
	osc fric relev	rain inc	3111003 (CITCIC	s your respect totale	110113:], 111011 ¢	nodoce a reson with the new r
	fun			() •	
Use the relevant methods (circle your helper functions!), then produce a result with the nev					/ •	<u>Define</u> :
Use the relevant methods (circle your helper functions!), then produce a result with the new fun(<u>/ </u>					 Are there mo
Use the relevant methods (circle your helper functions!), then produce a result with the new fun ():						
Use the relevant methods (circle your helper functions!), then produce a result with the new fun ():						Are there
Use the relevant methods (circle your helper functions!), then produce a result with the new fun ():						Are there t
† = Define Are there Are there						Are there t

Bad Sample Tables!

For each word problem, a Sample Table must have (1) all the columns that matter, (2) a representative sample of the rows, and be in (3) random order. For each problem below, check the boxes to determine if the Sample Table meets those criteria.

١.	The shelter wan	ts to know the	e median age	of all the cats

name	species	age	fixed	legs	pounds	weeks	Relevant columns
Sasha	cat	1	FALSE	4	6.5	3	Representative sample of rows
Mittens	cat	2	TRUE	4	7.4	5	Random order
Sunfower	cat	5	TRUE	4	8.1	10	

2. The shelter wants a pie chart showing all the dogs' weight

name	species	age
Fritz	dog	4
Wade	cat	2
Nibblet	rabbit	6
Daisy	dog	5

3. Sort all the animals alphabetically by name

name	species	age	fixed	legs	pounds	weeks	Delevered a alcuser
Ada	dog	2	TRUE	4	32	3	Relevant columnsRepresentative sample of rows
Во	dog	4	TRUE	4	76.1	10	□ Representative sample of fows □ Random order
Boo-boo	dog	11	TRUE	4	123	10	Random order

4. Make a bar chart for all the fixed animals

name	species	age	fixed	legs	pounds	weeks	П	Relevant columns
Sasha								Representative sample of rows
								Random order

Define a function bar-kitten-adoption, which takes in a Table of animals and creates a bar chart showing how many weeks it took for each kitten to be adopted

Contract and Purpos	e	
	•	\rightarrow
Examples		
Make a Start Table and	d a result based on that table.	
	_	
	<u> </u>	
Define the function		
	ods (circle your helper functions!),	then produce a result with the new table.
fun)	: Define the table
<u>† =</u>		
		Are there more columns?
		Are there fewer rows?
		Are the rows ordered?
		Produce the result
end		

Contract and Purpose			
	:		\rightarrow
Examples			
Make a Start Table and a re	esult based on that ta	ble.	
		→	
Define the function Use the relevant methods (circle your helper fund	ctions!), then produ	uce a result with the new table.
,	,	, , , , , , , , , , , , , , , , , , , ,	
fun	():	
			<u>Define the table</u>
, -			Are there more columns?
	 		 Are there fewer rows?
			Are the rows ordered?
			 Produce the result
end			

Contract and Purp	ose				
	:			-	
Examples					
Make a Start Table	and a result based	on that tabl	e.		
			_		
					
Define the functio					
Use the relevant me	thods (circle your l	helper functi	ons!), then pro	oduce a result	with the new table.
fun		() :		
_					Define the table
					Are there more columns?
					Are there fewer rows?
					Are the rows ordered?
					Produce the result
end					

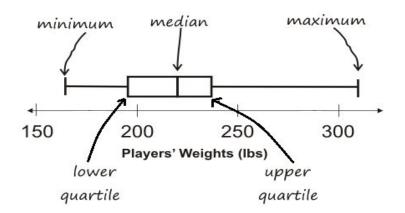
Visualizing My Dataset

What quantity charts did you make, and what do you notice? Fill in the table below.

Based on a chart of	I notice that

Unit 5

- There are three ways to measure the "center" of a dataset, to talk about a whole column of data using just one number:
 - 1. The **mean** of a dataset is the average of all the numbers
 - 2. The **median** of a dataset is a value that is smaller than half the dataset, and larger than the other half
 - 3. The **modes** of a dataset are the numbers that appear the most often.
- Data Scientists can also measure the "variation" of a dataset using a **five number summary:**
 - 1. The **minimum** the smallest value in the dataset
 - 2. The **first**, **or "lower" quartile (Q1)** the median value that separates the first quarter of the values in the dataset from the second quarter
 - 3. The **second quartile (Q2)** the median value which separates the entire dataset into "top" and "bottom" halves.
 - 4. The **third**, **or** "**upper**" **quartile (Q3)** the median value that separates the third quarter of the values in the dataset from the fourth quarter
 - 5. The **maximum** the largest value in the dataset
- The five number summary can be used to draw a box-and-whisker plot.



Summarizing Columns in Animals

The Column I choose to measure is					
Measures of Center The three measures for this column are:					
Mean (Ave	an (Average) Median N		Mode(s)		
Based on the diffe	erences betwee	en mean and medi	an, I conclude :		
Measures of Variation My five-number summary is:					
Minimum	Q1	Q2 (Median)	Q3	Maximum	
A box plot can be drawn from this summary on the number line below:					
From this summar	y and box-plot,	I conclude:			

The shelter wants a summary of the variation in ages among the dogs. Write a function called variation-dog-age that will take in a table of animals produce a box-plot that shows this variation.

Contract	and Purpo	ose																	
			:										\rightarrow						
													-						
	-																		
Example Make a St	s art Table a	nd a	result b	asec	on the	at tab	le.												
animals	s-table						\rightarrow	,	ar.	iat	ior	ı–do	og-	age	e(a	nim	als-	-tał	ole)
name	species	age	fixed	legs	weight	adopt		_							•				
Snowcone	cat	2	TRUE	4	6.1	5													
Lucky	dog	3	TRUE	3	45.4	9													
Hercules	cat	3	FALSE	4	13.4	7													
Toggle	dog	3	TRUE	4	48	3													
Snuggles	tarantula	2	FALSE	8	0.1	1													
	e function		/ • 1				•	1)							• • • •				
Use the re	levant met	noas	(circle	your	neiper	funct	ions	S!),	ther	n pro	duc	e a	resu	JIT W	ith t	ne n	ew to	.eldk	
					,			,											
					_ (_)	:							Def	ine t	he t	able
_ <i>bu</i>	<u>ild-column</u>	5(_)			e mor		
fil	ter()	1	re th	iere f	ewer	rows?
ord	der-by()	F	re th	ne row	's ord	ered?
														_	Pr	odu	ce th	ne re	esult
end																			

Interpreting Variation

Consider the following list dataset, representing the annual income of ten people:

\$65k, \$12k, \$14k, \$280k, \$15k, \$22k, \$45k, \$34k, \$45k, \$175k

1. In the space below, rewrite this dataset in **sorted order**.

2. In the table below, compute the **measures of center** for this dataset.

Mean (Average)	Median	Mode(s)

3. In the table below, compute the **five number summary** of this dataset.

Minimum	Q1	Q2 (Median)	Q3	Maximum

4. On the number line below, draw a **box plot** for this dataset.

5. The following statements are correct...but misleading. Write down the reason why.

Statement Why it's misleading

"The average person makes more than \$70k dollars!"

"The most common salary is \$45k/yr!"

"This group is really diverse, with people making as little as 12k and as much as \$280k!"

Summarizing a Column in My Dataset

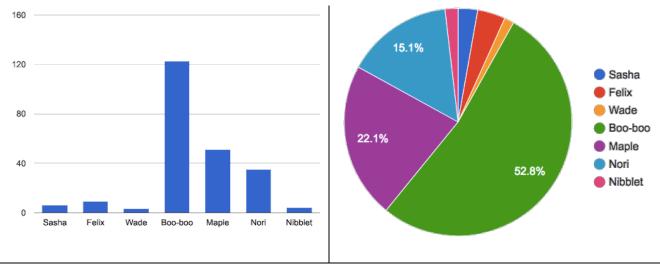
The column I choose to measure is						
Measures of Center The three measures for this column are:						
Mean (Ave	rage)	Median	٨	Mode(s)		
Based on the differences between mean and median, I conclude :						
		easures of Variat				
Minimum	Q1	Q2 (Median)	Q3	Maximum		
A box plot can be drawn from this summary on the number line below:						
From this summary and box-plot, I conclude:						

Unit 6

- Frequency Bar charts show the number of rows belonging to a given category. The more rows in each category, the longer the bar. Frequency bar charts provide a visual representation of the frequency of values in a categorical column. Since categorical data cannot be ordered, there is no strict ordering of bars in a frequency bar chart.
- Histograms show the number of rows that fall within certain ranges, or "bins" of a
 dataset. The more rows that that fall within a particular "bin", the longer the bar.
 Histograms provide a visual representation of the frequency of values in a
 quantitative column. Quantitative data can be ordered, so the bars of a
 histogram are always sorted.
- 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 distribution in the dataset.

Visualizing Quantity (Review)

Use the charts below to help you answer the questions.



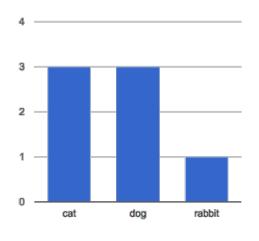
Animals Weights (lbs)

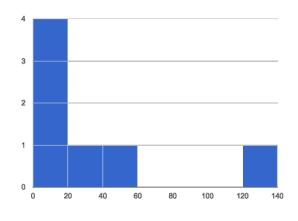
- 2. Which animal is the lightest?
- 3. How much of the total weight comes from Maple?
- 4. How much of the total weight comes from Nori?
- 5. Which chart did you use for questions 1 and 2?
- 6. Which chart did you use for questions 3 and 4?
- 7. Why are some questions easier to answer with one kind of chart or another?

Visualizing Frequency

name	species	age	pounds
"Sasha"	"cat"	1	6.5
"Boo-boo"	"dog"	11	123
"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?
- 2. How many dogs are there?
- 3. How many animals are between 3-6 years old?
- 4. How many weigh between 0-5 pounds?
- 5. Are there more animals weighing 0-5 than 6-10 pounds?
- 6. The charts below are based on the Sample Table above. What is each one measuring? Write down your guess underneath each one.





42

Define a function freq-bar-gender, which takes in a Table of animals and creates a frequency bar chart showing how many animals are male v. female.

Contract and Purpose				
	:		<i>></i> _	
Examples				
Make a Start Table and	a result based on	that table.		
		\rightarrow		
Define the function Use the relevant method	ds (circle your bels	por functional) ther	produce a regult	with the new table
ose me relevant memor	us (Circle your rielp	per forictions: j, men	i produce a resuir	wiiii iiie iiew labie.
fun	():		
<i>t</i> =				Define the table
				Are there more columns?
				Are there fewer rows?
				Are the rows ordered?
				Produce the result
end				

Define a function histogram-adoption, which takes in a Table of animals and creates a histogram showing how long it took for animals to get adopted

Contract and Purpose			
	:		\rightarrow
Examples			
Make a Start Table and a re	esult based on that tal	ole.	
		→	
Define the function		·	
Use the relevant methods (circle your helper fund	ctions!), then proc	duce a result with the new table.
	,	\	
	(Define the table
<u>† =</u>			 Are there more columns?
			
			Are there fewer rows?
			Are the rows ordered?
			Produce the result
end			

Visualizing My Dataset

What frequency charts did you make, and what do you notice? Fill in the table below.

Based on a chart of	I notice that

Matching Charts to Questions

For each of the questions below, draw a line to the chart that will best answer it.

als at the shelter	Are there more of the animals at the shelter fixed or unfixed?	1.
cat wait to be	How many weeks did each cat wait to be adopted?	2.
dogs are there? Bar Chart	How many male v. female dogs are there?	3.
	How many animals have 4 legs? 8? 3?	4.
	What percent of the total weight at the shelter is made up by Boo-boo?	5.
eights across all	What is the distribution of weights across all the animals older than 3?	6.
re of each Histogram	How many animals are there of each species?	7.
be adopted?	Who waited the longest to be adopted?	8.

Unit 7

- **Scatter Plots** show the 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 pattern, it is possible that a relationship or **correlation** exists between those two columns.
- If there is a pattern to the points in a scatter plot, points that are far away from the pattern are called **outliers**.
- We can express this correlation by drawing line through the data cloud, so that
 the distance between the line and each of the points is as small as possible. This
 line is called the line of best fit or predictor function and allows us to make
 predictions based on the dataset.

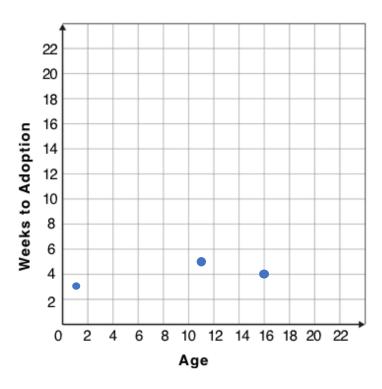
(Dis)Proving a Claim

"Younger animals are cuter, so they get adopted faster."

Do you agree? If so, why?
I hypothesize
What would you look for in the dataset to see if you are right?

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



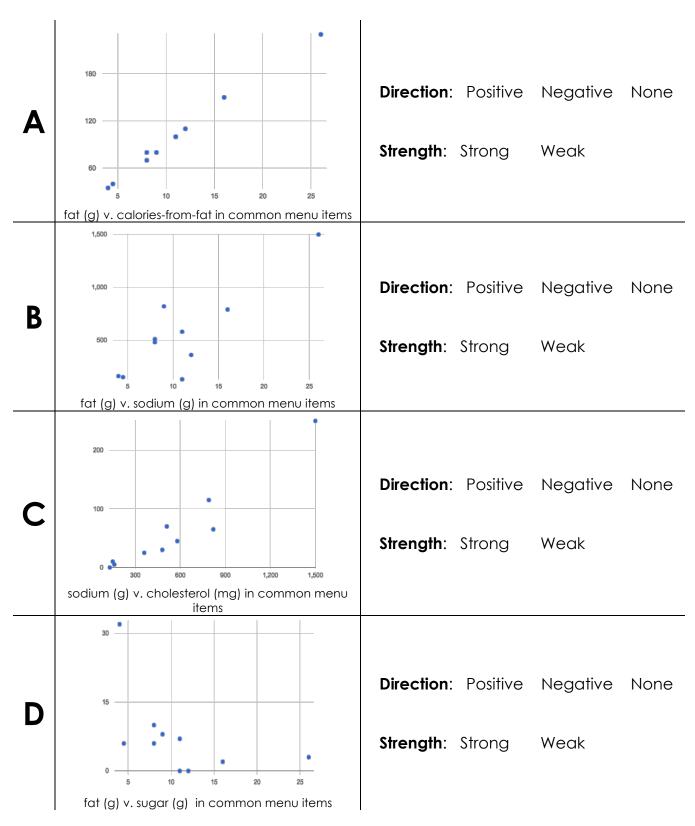
- 1. For each row in the Sample Table on the left, add a point to the scatter plot on the right. The first 3 rows have been completed for you. Use the values from the age column for the x-axis, and values from the weeks column for the y-axis.
- 2. Do you see a pattern? Do the points seem to shift up or down as age increases? **Draw a line on the scatter plot to show this pattern**.
- 3. Does the line slope upwards or downwards?
- 4. Are the points mostly close to the line?

Define a function <code>dogs-age-weeks</code>, which takes in a Table of animals and creates a scatter plot of all the dogs, tracking their <code>age</code> on the x-axis and the number of <code>weeks</code> it took for them to be adopted on the y-axis.

Contract and Purpose				
	:		-	
Evernolos				
Examples Make a Start Table and c	result based on that	table.		
		_ →		
		-		
		-		
		-		
		-		
		J		
Define the function				
Use the relevant methods	(circle your helper fu	unctions!), then	produce a result	with the new table.
5	,	\		
fun <u>† =</u>	(Define the table
7				Are there more columns?
				Are there fewer rows?
				Are the rows ordered?
				Produce the result
end				

Drawing Predictors

For each of the scatter plots below, draw a **predictor line** that fits best.



Correlations in My Dataset

1) There may be a correlation between _		and
	column	
I think it is a		
COLUMN	strong / weak	positive / negative
correlation, because		
correlation, because		
	It would	d be stronger if I looked
		_
a subset	or extension of my da	n+ a
a subsec	or excension or my da	i ca
1) There may be a correlation between _		and
	column	
I think it is a		
column	strong / weak	positive / negative
correlation because		
correlation, because		
	. It would	d be stronger if Hooked
	, ,, oo.,	
a subset		
a subset	or extension of my da	ita
1) There may be a correlation between		and
Ty more may be a conclamon between _	column	and
I think it is a		
column	strong / weak	positive / negative
correlation, because		
	It would	d he stronger if Llooked
		a be shoriger in Hooked
a subset		
a subset	or extension of my da	ıta

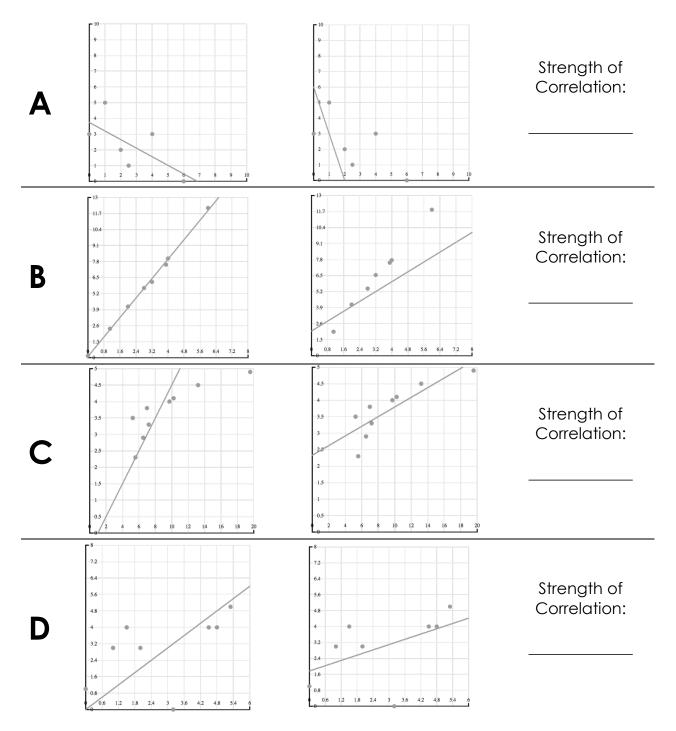
Unit 8

- Given a **predictor function** and a scatter plot, we can compute the error by adding the squares of all the distances between the function and each point in the plot. The error is called the **r**² **statistic**, which tells us how much of the variation in the y-axis can be explained by the x-axis.
- A strong correlation will have a large r². A weak correlation will have a small r².
- A **positive correlation** means the slope of the line of best fit is positive. A **negative correlation** means the slope is negative.
- **Linear Regression** is a way of computing the **line of best fit**, by taking a scatter plot and deriving the slope and y-intercept for a line that has the smallest possible r².
- <u>Correlation is not causation!</u> Correlation only suggests that two measures are related, but does not tell us if one causes the other. For example, hot days are correlated with people running their air conditioners, air conditioners do not cause hot days!

Grading Predictors

Below are the scatter plots for data sets A-D, with two different lines predictor lines drawn on top. For plots A-D:

- 1. Circle the plot with the line that fits better
- 2. Give the plot you circled a grade between 0 (no correlation) and 1 (perfect correlation)



Findings in the animals Dataset

I performed a linear	r regression on	do	ogs at the shelter	, and
			dataset or subset	
found	a weak (r²=	0.25), positive	itive/negative correlation k	petween
age of the doc	gs (in weeks)	and number	er of weeks to be adopted	From this, I
	25% of the		[y-axis] doption time is explained is explained by [x-axis]	
		ation in [y-axis]	is explained by [x-axis]	
by the age of th	e dog			
I performed a linear	r regression on		dataset or subset	, and
•	_		dataset or subset	
found			correlation k	oetween
		and	[y-axis]	From this, I
[x-a	axis]		[y-axis]	
conclude that	7,		is explained by [x-axis]	
	r' % of the varia	ation in [y-axis]	is explained by [x-axis]	
				·
I performed a linear	r rearession on			, and
,			dataset or subset	
found			correlation k	oetween
	a strong,	/weak (r^2 =), pos	correlation k	
		and		. From this, I
[x-a	axis]		[y-axis]	
conclude that				
<u></u>	r^2 % of the varia	ation in [y-axis]	is explained by [x-axis]	
				_
				·

Correlations in My Dataset

I performed a linear regression on			, and
		dataset or subset	
founda strong/w		correlation	between
a strong/w	eak $(r^2 =)$, pos	sitive/negative	
	_ and	[y-axis]	From this, I
[x-axis]		[y-axis]	
conclude that $\underline{\hspace{2cm}}$ r² % of the variat			
r^2 % of the variat	ion in [y-axis] is explained by [x-axis]	
			•
I performed a linear regression on			, and
		dataset or subset	
founda strong/w		correlation	between
a strong/w	eak $(r^2 =)$, pos	sitive/negative	
	_ and	[y-axis]	From this, I
[x-axis]		[y-axis]	
conclude that $\underline{\hspace{1cm}}$ r² % of the variat			
r² % of the variat	ion in [y-axis] is explained by [x-axis]	
I performed a linear regression on			, and
		dataset or subset	
found	2	sitive/negative correlation	between
a strong/w	eak (r ² =), po:	sitive/negative	
	_ and	[y-axis]	From this, I
[x-axis]		[y-axis]	
conclude that $\underline{\hspace{1cm}}$ r² % of the variat			
r^- % of the variat	ion in [y-axis] is explained by [x-axis]	
			·

Unit 9

Fake News!

Every claim below is wrong! Your job is to figure out why, by looking at the data.

	Data	Claim	Why it's wrong
1	The average player on a basketball team is 6'1".	"Most of the players on the team are taller than 6'."	Wily it's wiong
2	After performing linear regression on census data, a positive correlation (r ² =0.18) was found between people's height and salary.	"Taller people get paid more."	
3	y=12.234x + -17.089; r-sq: 0.636	"According to the predictor function indicated here, the value on the x-axis is will predict the value on the y-axis 63.6% of the time."	
4	15 Sasha Felix Wade Boo-boo Maple Nori Bar Chart of Pet Ages	"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	20 40 60 80 100 120 140 160 180 Weight (pounds)	"According to this histogram, most animals weigh between 40 and 60 pounds."	
6	After performing linear regression, a negative correlation (r ² =0.91) was found between the number of hairs on a person's head and their likelihood of owning a wig.	"Owning wigs causes people to go bald."	

Blank Recipes, Table Plans, and References

Design Recipes

	:		\rightarrow	
naı	me	domain		range
 xamples	:			
	() is		
	(
nd	(, 15		
un	() :		
nd				
	:			
naı	me	domain		range
xamples	:			
	() is		
	(
nd				
un	() :		
nd				

Design Recipes

	•		\rightarrow	
name		domain		range
ples:				
	() is		
	(, is		
	() :		
			\rightarrow	
	:		⁷ _	
name	•	domain	~ _	range
name	•	domain	7	range
	•	domain	7	range
	•	domain	7 <u>_</u>	range
	•		7	range
	(domain) is	7	range
	() is		
	(
	() is		
ples:	(() is		
name ples:	(() is		

Design Recipes

	•		\rightarrow	
name		domain		range
ples:				
	() is		
	(, is		
	() :		
			\rightarrow	
	:		⁷ _	
name	•	domain	~ _	range
name	•	domain	7	range
	•	domain	7	range
	•	domain	7 <u>_</u>	range
	•		7	range
	(domain) is	7	range
	() is		
	(
	() is		
ples:	(() is		
name ples:	(() is		

Contract and Purpose			
	:		\rightarrow
_			
Examples Make a Start Table and a re	sult based on that to	ble	
make a diam rable and a re	son basea on marra		
		\rightarrow	
Define the function			
	circle your helper fund	ctions!), then prod	uce a result with the new table.
,	, ,	, ,	
fun	():	
			Define the table
<u>† =</u>			 Are there more columns?
			 Are there fewer rows?
			·
			Are the rows ordered?
			Produce the result
end			

Contract and Purpose			
	:		\rightarrow
			
Examples Make a Start Table and a	result based on that tal	ole.	
THERE GIGHT TEETS GITE G	roson basea en mar ia.		
		\rightarrow	
Define the function			
	(circle your helper func	tions!), then produ	uce a result with the new table.
fun	():	
			<u>Define the table</u>
·			Are there more columns?
			 Are there fewer rows?
			 Are the rows ordered?
			<u>Produce the result</u>
end			

Contract and Purpose			
	_:		_
Examples			
Make a Start Table and a re	esult based on that to	able.	
		→	
Define the function Use the relevant methods (circle your helper fur	nctions!), then produc	e a result with the new table.
fun	():	Define the table
<u>† =</u>			
			Are there more columns? Are there fewer rows?
			Are the rows ordered?
 end			<u>Produce the result</u>

Contracts

Name	Domain		Range
triangle	:: (side :: Number, style :: String, color :: String)	\rightarrow	Image
circle	:: (radius :: Number, style :: String, color :: String)	\rightarrow	Image
star	:: (radius :: Number, style :: String, color :: String)	\rightarrow	Image
rectangle	:: (width :: Num, height :: Num, style :: Str, color :: Str)	\rightarrow	Image
ellipse	:: (width :: Num, height :: Num, style :: Str, color :: Str)	\rightarrow	Image
square	:: (size :: Number, style :: String, color :: String)	\rightarrow	Image
text	:: (str :: String, size :: Number, color :: String)	\rightarrow	Image
overlay	:: (img1 :: <i>Image</i> , img2 :: <i>Image</i>)	\rightarrow	Image
rotate	:: (degree :: Number, img :: Image)	\rightarrow	Image
scale	:: (factor :: Number, img :: Image)	\rightarrow	Image
string-repeat	:: (text :: String, repeat :: Number)	\rightarrow	String
string-contains	:: (text :: String, search-for :: String)	\rightarrow	Boolean
num-sqr	:: (n :: Number)	\rightarrow	Number
num-sqrt	:: (n :: Number)	\rightarrow	Number
num-min	:: (a :: Number, b:: Number)	\rightarrow	Number
num-max	:: (a :: Number, b:: Number)	\rightarrow	Number
get-row	:: (t :: Table, index :: Number)	\rightarrow	Row

Contracts

Name	Domain		Range
<table>.row-n</table>	:: (n :: Number)	\rightarrow	Row
<table>.filter</table>	:: (test :: (Row → Boolean))	\rightarrow	Table
<table>.build-column</table>	:: (col :: String, builder :: (Row → Value))	\rightarrow	Table
mean	:: (<u>t</u> :: Table, col :: String)	\rightarrow	Number
median	:: (t :: Table, col :: String)	\rightarrow	Number
modes	:: (t :: Table, col :: String)	\rightarrow	List <number></number>
bar-chart	:: (t :: Table, labels :: String, values :: String)	\rightarrow	Image
pie-chart	:: (t :: Table, labels :: String, values :: String)	\rightarrow	Image
box-plot	:: (t :: Table, col:: String)	\rightarrow	Image
freq-bar-chart	:: (t :: Table, values :: String)	\rightarrow	Image
histogram	:: (t :: Table, values :: String, bin-width :: Number)	\rightarrow	Image
scatter-plot	:: (t :: Table, xs :: String, ys :: String)	\rightarrow	Image
labeled-scatter-plot	:: (t :: Table, labels :: String, xs :: String, ys :: String)	\rightarrow	Image
lr-plot	:: (t :: Table, xs :: String, ys :: String)	\rightarrow	Image
labeled-lr-plot	:: (t :: Table, labels :: String, xs :: String, ys :: String)	\rightarrow	Image