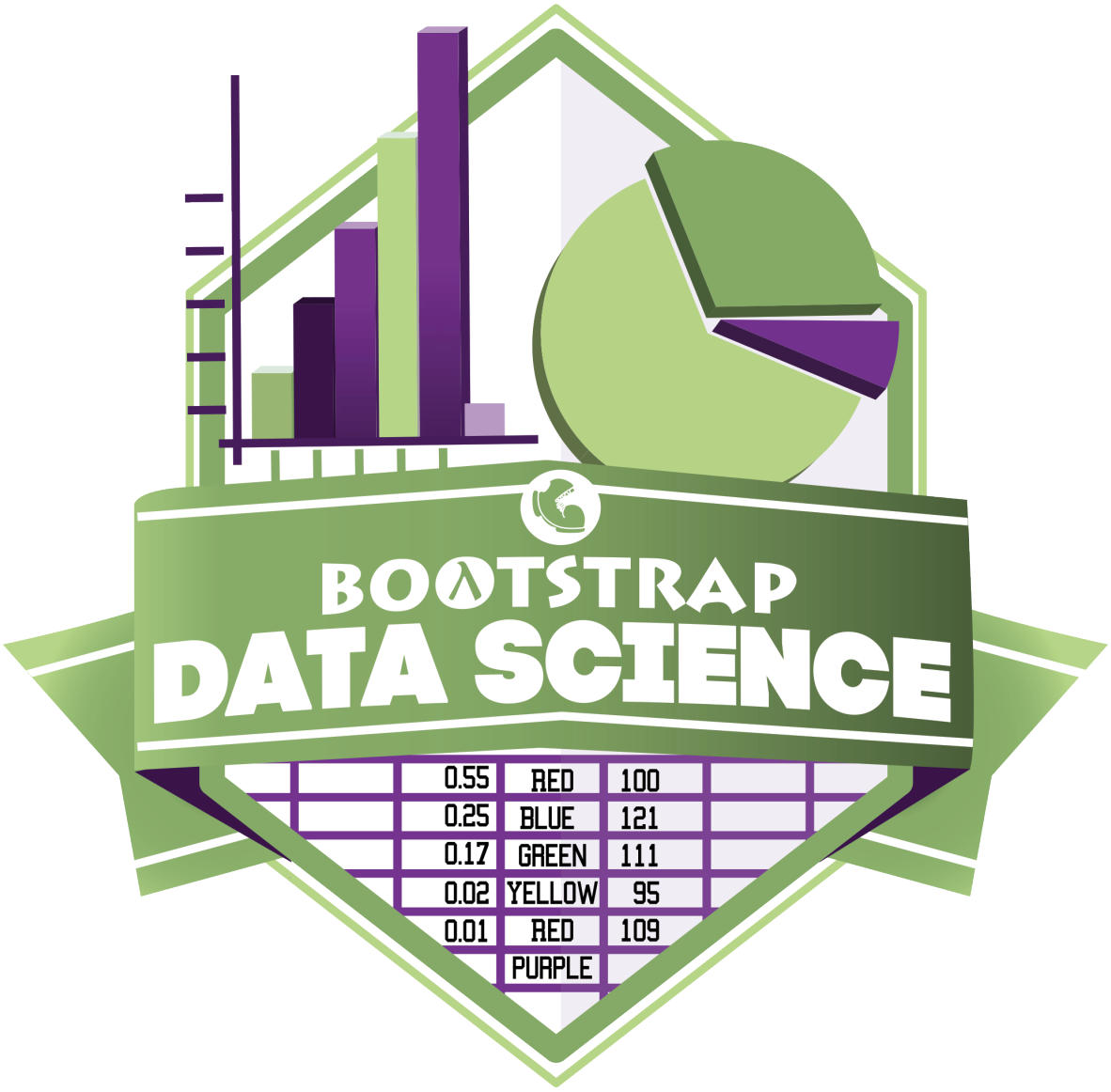


BOOTSTRAP
www.bootstrapworld.org

Workbook v1.1

Brought to you by the Bootstrap team:

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

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")`

- **Examples** help programmers reason about their code. Every example contains two expressions, and the example "passes" if both expressions evaluate to the same thing. For example: `4 + 2 is 6`, or `"cat" == "dog" is false`

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!). What happens when you hit "Enter"?

4. Try typing your name with the opening quote, but *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:

Operators

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 will always evaluate to either `true` ("yes") or `false` ("no"). What will each of the expressions below evaluate to? Write down the result in the blanks provided, and 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!

```
(3 <= 4) and (3 == 2)
```
_____

```
("a" == "b") and (3 <> 4)
```
_____

```
(3 <= 4) or (3 == 2)
```
_____

```
("a" == "b") or (3 <> 4)
```
_____

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

Playing with Tables

The table below represents four animals at the shelter:

animals

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

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

`get-row(animals, 3)`	Evaluates to `51.6`
`get-row(animals, 0)`	Evaluates to the last row in the table
`get-row(animals, 1)["gender"]`	Evaluates to `"male"`
`get-row(animals, 2)["age"]`	Evaluates to `"Toggle"`
`get-row(animals, 1)["weight"]`	Evaluates to `92`
`get-row(animals, 0)["name"]`	Evaluates to `6`
`get-row(animals, 3)["weight"]`	Evaluates to `female`
`get-row(animals, 2)["gender"]`	Evaluates to the first row in the table

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

a. *get-row(animals, 3)["name"]*	`"Maple"`
b.	`male`
c.	`4`
d.	`48`
e.	`"Nori"`

Writing Examples

1. In the examples block below, **put an "X" next to the examples that will *fail*.** Remember: examples only pass if the left- and right-hand expressions evaluate to the same thing!

```
examples:
  1 + 2 + 9                      is 19
  num-sqrt(16)                   is 2 + 2
  3 > 99                         is true
  square(10, "solid", "red")     is rectangle(10, 10, "solid", "red")
end
```

2. In the examples block below, **fill in the blank on the right-hand side so the example will *pass*.**

```
examples:
  string-repeat("yeah! ", 3)      is _____________________________________

  string-contains("Maya", "May")  is _____________________________________

  "apples" <> "oranges"           is _____________________________________

end
```

3. The examples block below refers to the `shapes` table on the right, using row-accessors and the `get-row` function. For each example, **fill in the blank so the example will *pass*.**

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

```
examples:
  get-row(_____________, 3)["name"]    is "square"

  get-row(shapes, _____)["corners"]    is 3

  get-row(shapes, 2)[_____________]    is true

end
```

Unit 2

Answering Questions from Data can take many forms. Here are a few types of questions, each requiring a different kind of analysis:

- **Lookup Questions** can be answered just by finding the right row and column a table. (e.g. – "How old is Toggle?")

- **Comparison Questions** can be answered by comparing a single row or column to all the rest of the table. (e.g. – "What is the heaviest animal at the shelter?")

- **Pattern Questions** require looking for trends across multiple rows or columns. (e.g. – "Do cats tend to be adopted sooner than dogs?")

Threats to Validity can undermine a conclusion, even if the analysis was done correctly. Some examples of threats are:

- **Sample size** – averaging the age of only three animals won't tell us anything reliable about the age of animals at the shelter!

- **Selection bias** – identifying the favorite food of the rabbits won't tell us anything reliable about what all the animals eat.

- **Sample error** – surveying dogs when they are puppies won't tell us anything reliable about overall dog behavior, since their behavior changes as they age.

- **Confounding variables** – if they person surveying the animals has a piece of bacon in their pocket, they will incorrectly find that all dogs are friendly!

The Animals Dataset

1. This dataset is Animals from an animal shelter

2. The columns are….
   (For each column in this dataset, fill out the datatype, and whether it contains Qualitative or Categorical data in the table below)

name	species	gender	age	fixed	pounds	weeks
"Sasha"	"cat"	"female"	1	false	6.5	3
"Toggle"	"dog"	"female"	3	true	48	1
"Boo-boo"	"dog"	"male"	11	true	123	24
"Snuggles"	"tarantula"	"female"	2	false	.01	1
"Nori"	"dog"	"female"	6	true	35.3	1
"Snowcone"	"cat"	"female"	2	true	6.1	5
Datatype						
Quantitative or Categorical?						

3. For the questions below, check the box next to the questions you COULD answer given this dataset:

   - ☐ **How old is Boo-boo?**
   - ☐ **What color is Snowcone's fur?**
   - ☐ **What is the average age of all the animals in the shelter?**
   - ☐ **Are there more fixed or unfixed animals at the shelter?**
   - ☐ **Are families with children more likely to adopt kittens?**

4. Some questions I have about this dataset:

What Questions Can You Answer?

The following is a dataset of a bicycle rider's training rides.

date	miles	time	weather	average speed	max speed
04/10/2018	10	44	"cloudy"	13	30
05/30/2018	15	66	"sunny"	13.5	22
06/12/2018	12	61	"rainy"	11.2	25
06/22/2018	15	61	"cloudy"	13	28
07/04/2018	24	103	"sunny"	14	26
07/12/2018	24	120	"windy"	12.5	26

For the questions below, check the box next to the questions you COULD answer given this dataset:

- ☐ **What is the cyclist's average speed across all rides?**
- ☐ **How many miles did they ride in June?**
- ☐ **What is the tallest hill this cyclist climbed?**
- ☐ **Does this cyclist ride slower when it is rainy?**
- ☐ **Does this cyclist ride faster when they are late to an appointment?**

For the questions below, check the box next to the questions you COULD NOT answer given this dataset:

- ☐ **What tire pressure produces the highest average speed?**
- ☐ **What is the average time it takes this cyclist to ride one mile?**
- ☐ **Does this cyclist ride more in April or July?**
- ☐ **What is the average temperature while this cyclist is riding?**
- ☐ **How many flat tires did this cyclist fix in June?**

Threats to Validity

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

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

The animal shelter noticed a large increase in pet adoptions between Thanksgiving and New Year's Eve. They conclude that at this current rate, there will be a huge demand for pets between January and April.

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

Threats to Validity

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 found that spider and rabbit food was by far the most popular cuisine!

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

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 then the others.

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

My Dataset

1. My dataset is _____

2. Some of my columns are….
   (Copy six columns from your dataset, and for each column write its datatype, and whether it contains Qualitative or Categorical data in the table below)

Datatype					
Quantitative or Categorical?					

3. Some questions I have about this dataset:

4. What are some possible threats to validity you may encounter in your analysis?

Unit 3

- 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 Design Recipe

Define a function called `is-fixed`, which tells us whether or not an animal is fixed

is-fixed :: *(animal :: Row)* → *Boolean*

name domain range

\# *Consumes an animal, and produces the value in the fixed column*

examples:

is-fixed (*sasha*) **is** *sasha["fixed"]*

_____ (_____) **is** _____

end

fun _____ (_____) **:** _____

end

Define a function called `gender`, which consumes a Row of the animals table tells us the gender of that animal

_____ :: _____ → _____

name domain range

\# _____

examples:

_____ (_____) **is** _____

_____ (_____) **is** _____

end

fun _____ (_____) **:** _____

end

Define a function called `is-cat`, which consumes a Row of the animals table and produces `true` if it's a cat.

is-cat :: *(animal :: Row)* → *Boolean*

name domain range

\# *Consumes an animal, and return true if the species is "cat"*

examples:

is-cat (*sasha*) **is** _____

_____ (_____) **is** _____

end

fun _____ (_____) **:** _____

end

Define a function called `is-young`, which consumes a Row of the animals table and produces `true` if it's an animal that is less than two years old.

_____ :: _____ → _____

name domain range

\# _____

examples:

_____ (_____) **is** _____

_____ (_____) **is** _____

end

fun _____ (_____) **:** _____

end

Define a function called `nametag`, prints out each animal's name in big red letters.

nametag :: *(animal :: Row)* → *Image*

name domain range

\# *Consumes an animal, and produces an image of their name in big, red letters*

examples:

nametag (*sasha*) **is** _____

_____ (_____) **is** _____

end

fun _____ (_____) **:** _____

end

Define a function called `is-kitten`, which consumes a Row of the animals table and produces true if it's a cat younger than two years old.

_____ :: _____ → _____

name domain range

\# _____

examples:

_____ (_____) **is** _____

_____ (_____) **is** _____

end

fun _____ (_____) **:** _____

end

Unit 4

- **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, `<table>.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:
  - <Table>.order-by – order the rows of a table based on a column
  - <Table>.filter – create a **subset** of the data, with only certain rows
  - <Table>.build-column – use the columns of a table to compute a new one

Reviewing Functions

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

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 the `Table` to which it is attached.

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`*

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.

Playing with Methods

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

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

The table **t** below represents four animals at the shelter:

name	gender	age	fixed	weight
“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).

```
t.order-by(“age”, true)
```

Produces a table containing *only* Toggle and Maple

```
t.filter(is-fixed)
```

Produces a table, sorted oldest-to-youngest.

```
t.build-column(“sticker”, nametag)
```

Produces a table, sorted youngest-to-oldest

```
t.filter(is-young)
```

Produces a table with an extra column, named “sticker”

```
t.order-by(“age”, false)
```

Produces a table containing Maple and Toggle, in that order.

```
t
 .filter(is-young)
 .order-by(“weight”, false)
```

Produces a table containing the same four animals.

```
t
 .order-by(“age”, true)
 .build-column(“sticker”, nametag)
```

Produces a table with an extra “sticker” column, sorted youngest-to-oldest

Unit 5

- Functions can contain value definitions
- We use **Table Plans** to help us use table methods correctly, without making mistakes

Review

- In the Interactions Area, use table methods to sort your table by one column. **Try sorting your table in both ascending and descending order.**

- If a researcher is looking at a dataset of students, they might want to divide the data into separate populations of boys and girls. A veterinarian might want to look at only the cats at a shelter. **Copy one of your "filtering" answers from Page 18 below**, to define the filtering criteria you want to use**.**

- In the space below, **use the Design Recipe to write a function that checks if a row in your dataset fits that criteria.** Whatever criteria you choose, it should be true for some rows and false for others. **Type this function into the Definitions Area.**

```
_________________::_________________________ → _________________
        name                  domain                    range
# ____________________________________________________________
examples:

    _____________(sample1) is ____________________________

    _____________(sample2) is ____________________________
end

fun _____________ (________) :  ______________________________

end
```

- **Use the function to filter your dataset.**

- Instead of using the function you wrote to *filter* your dataset, **use another table method to build a new column** that shows whether or not each row meets the criteria.

Table Plan

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.

Contract and Purpose

get-kittens-tags **::** *(animals :: Table)* → *Table*

 # Consume a table of animals, and produce a table containing kittens with nametags, sorted by name

Example Tables

Make a Start Table and a result based on that table.

`animals-table` → `get-kittens-tags(animals-table)`

name	species	age	fixed	legs	weight	adopt
Sasha	cat	1	FALSE	4	6.5	4
Toggle	dog	3	TRUE	4	48	3
Buddy	lizard	2	FALSE	4	0.3	12
Wade	cat	1	FALSE	4	3.2	4
Mittens	cat	2	TRUE	4	7.4	5

name	species	age	fixed	legs	weight	adopt	tag
Sasha	cat	1	FALSE	4	6.5	4	Sasha
Wade	cat	1	FALSE	4	3.2	4	Wade

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

```
fun get-kittens-tags ( animals ) :
    t = animals                                  Define the table
    .build-column(                              )  Are there more columns?
    .filter(                                    )  Are there fewer rows?
    .order-by(                                  )  Are the rows ordered?
    t                                            Produce the result
end
```

Table Plan

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 Purpose

get-dogs-by-age **::** *(animals :: Table)* → *Table*

 # Consume a table of animals, and produce a table containing only the dogs, sorted by age

Examples

Make a Start Table and a result based on that table.

`animals-table` → `get-dog-by-age(animals-table)`

name	species	age	fixed	legs	weight	adopt
Snowcone	cat	2	TRUE	4	6.1	5
Wade	cat	1	FALSE	4	3.2	4
Hercules	cat	3	FALSE	4	13.4	7
Toggle	dog	3	TRUE	4	48	3
Fritz	dog	4	TRUE	4	92	6

name	species	age	fixed	legs	weight	adopt
Toggle	dog	3	TRUE	4	48	3
Fritz	dog	4	TRUE	4	92	6

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

```
fun ____________________ (_________):
    t = ____________________________________________
    .build-column(____________________________________)
    .filter(_________________________________________)
    .order-by(_______________________________________)
    ________________________________________________
end
```

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

Table Plan

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!

Contract and Purpose

_____ : : _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

`animals-table` → `old-dogs-diet(animals-table)`

name	species	age	fixed	legs	weight	adopt
Snowcone	cat	2	TRUE	4	6.1	5
Lucky	dog	3	TRUE	3	45.4	9
Mr. PB	dog	10	FALSE	4	161	6
Boo-boo	dog	11	TRUE	4	123	24
Snuggles	tarantula	2	FALSE	8	0.1	1

name	species	age	fixed	legs	weight	adopt
Mr. PB	dog	10	FALSE	4	161	6
Boo-boo	dog	11	TRUE	4	123	24

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) :

t = _____ *Define the table*

.build-column(_____ *)* *Are there more columns?*

.filter(_____ *)* *Are there fewer rows?*

.order-by(_____ *)* *Are the rows ordered?*

_____ *Produce the result*

end

Table Plan

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 Purpose

_____ :: _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

`animals-table` → `get-fixed-by-legs(animals-table)`

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

name	species	age	fixed	legs	weight	adopt	year
Snowcone	cat	2	TRUE	4	6.1	5	2015
Lucky	dog	3	TRUE	3	45.4	9	2014
Toggle	dog	3	TRUE	4	48	3	2014

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

```
fun _____________________ (_________):
    t = ____________________________________________
    __.build-column(____________________________________)
    __.filter(_________________________________________)
    __.order-by(_______________________________________)
    ____________________________________________________
end
```

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

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 6

- **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:
  - At least all the relevant columns
  - Enough rows to accurately represent the dataset
  - Rows that are randomly-ordered

Statements about Columns

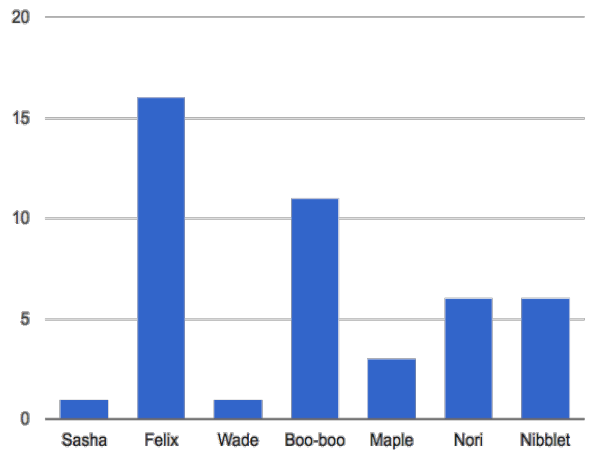
Use the Table below to help you answer the questions.

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

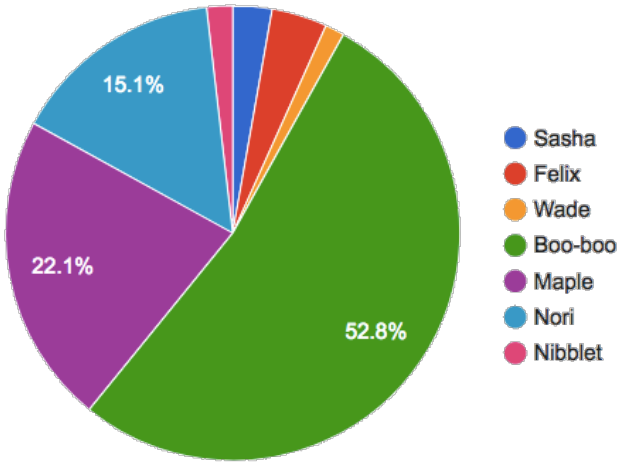
1. Which animal(s) is/are the heaviest? _____
2. Which animal(s) is/are the youngest? _____
3. How much of the *total weight* comes from Maple? _____
4. How much of the *combined age* comes from Nori? _____
5. Would these questions be harder to answer if the table had 100 rows? If so, why?

Visualizing Quantity

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



Animals Ages (yrs)



Animals Weights (lbs)

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	

Table Plan

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.

Contract and Purpose

_____ **::** _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

`animals-table` → `pie-dog-weight(animals-table)`

name	...	weight
Snowcone	…	6.1
Lucky	…	45.4
Hercules	…	13.4
Toggle	…	48
Snuggles	…	0.1

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) **:**

t = _____

_____

_____

_____

_____

end

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

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.

1. The shelter wants to know the median age of all the cats

name	species	age	fixed	legs	pounds	weeks
Sasha	cat	1	FALSE	4	6.5	3
Mittens	cat	2	TRUE	4	7.4	5
Sunfower	cat	5	TRUE	4	8.1	10

- ☐ Relevant columns
- ☐ Representative sample of rows
- ☐ Random order

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

- ☐ Relevant columns
- ☐ Representative sample of rows
- ☐ Random order

3. Sort all the animals alphabetically by name

name	species	age	fixed	legs	pounds	weeks
Ada	dog	2	TRUE	4	32	3
Bo	dog	4	TRUE	4	76.1	10
Boo-boo	dog	11	TRUE	4	123	10

- ☐ Relevant columns
- ☐ Representative sample of rows
- ☐ Random order

4. Make a bar chart for all the fixed animals

name	species	age	fixed	legs	pounds	weeks
Sasha	cat	1	FALSE	4	6.5	3

- ☐ Relevant columns
- ☐ Representative sample of rows
- ☐ Random order

Table Plan

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 Purpose

_____ **::** _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

_____ → _____

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) **:**

t = _____

_____

_____

_____

_____

end

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

Table Plan

Contract and Purpose

_____ : : _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

_____ → _____

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) :

t = _____

_____

_____

_____

_____

end

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

Table Plan

Contract and Purpose

_____ **::** _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

_____ → _____

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) **:**

t = _____

_____

_____

_____

_____

end

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

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 7

- 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

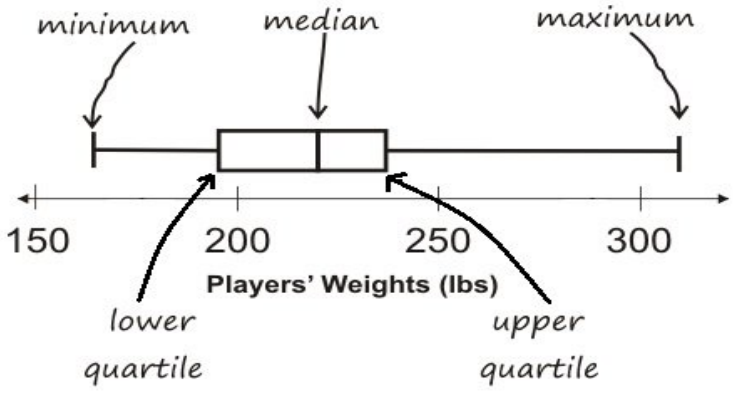
    2. The **first, or "lower" quartile (Q1)** – the median value that separates the first quarter of the values in the dataset from the second quarter

    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 ***weeks***

Measures of Center

The three measures for this column are:

Mean (Average)	Median	Mode(s)

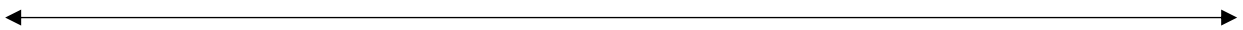
Based on the differences between mean and median, I conclude :

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 summary and box-plot, I conclude:

Table Plan

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 and produce a box-plot that shows this variation.

Contract and Purpose

_____ : : _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

`animals-table` → `variation-dog-age(animals-table)`

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

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

```
fun _____ (_____):
  t = _____                                  Define the table
  .build-column(_____)                       Are there more columns?
  .filter(_____)                             Are there fewer rows?
  .order-by(_____)                           Are the rows ordered?
  _____                                      Produce the result
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**.

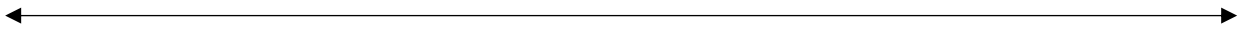
2. In the table below, compute the **measures of center** for this dataset.

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
"They're rich! The average person makes more than $70k dollars!"	
"It's a middle-income list: 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 (Average)	Median	Mode(s)

Based on the differences between mean and median, 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 summary and box-plot, I conclude:

Unit 8

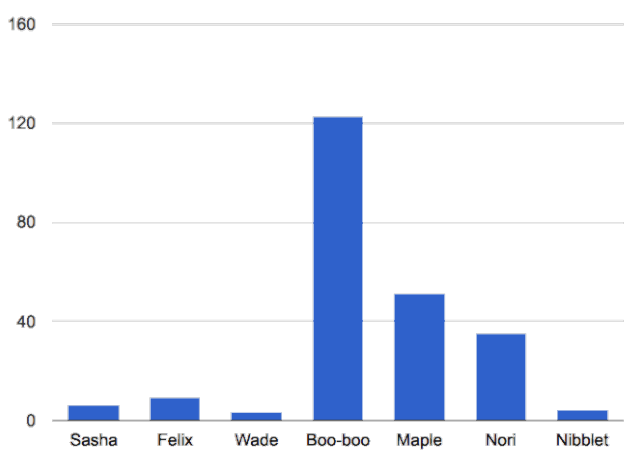
- **Frequency Bar charts** show the number of rows belonging to a given category. The more rows in each category, the longer the bar. Frequency bar charts provide a visual representation of the frequency of values in a **categorical** column. Since categorical data cannot be ordered, there is no strict ordering of bars in a frequency bar chart.

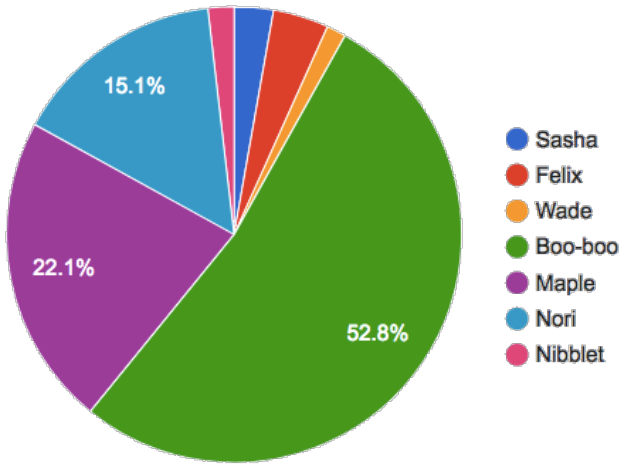
- **Histograms** show the number of rows that fall within certain ranges, or "bins" of a dataset. The more rows that that fall within a particular "bin", the longer the bar. Histograms provide a visual representation of the frequency of values in a **quantitative** column. Quantitative data can be ordered, 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.





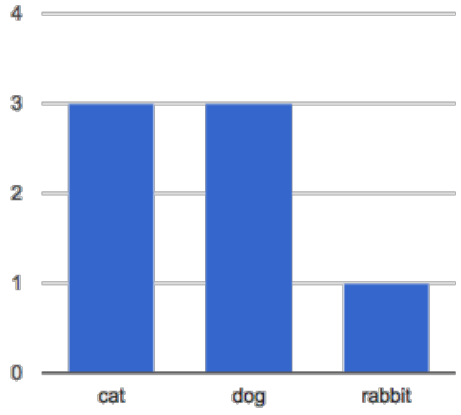
Animals Weights (lbs)

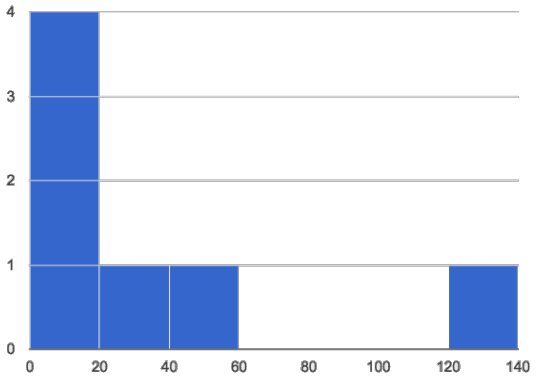
1. Which animal is the heaviest? _____
2. Which animal is the lightest? _____
3. How much of the *total weight* comes from Maple? _____
4. How much of the *total weight* comes from Nori? _____
5. Which chart did you use for questions 1 and 2? _____
6. Which chart did you use for questions 3 and 4? _____
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.





_____ _____

_____ _____

_____ _____

_____ _____

Table Plan

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

_____ **::** _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

_____ → _____

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) **:**

t = _____

_____

_____

_____

_____

end

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

Table Plan

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

_____ **::** _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

_____ → _____

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) **:**

t = _____

_____

_____

_____

_____

end

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

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. (You may find that more than one question is best answered by the same chart!)

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

Unit 9

- **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."

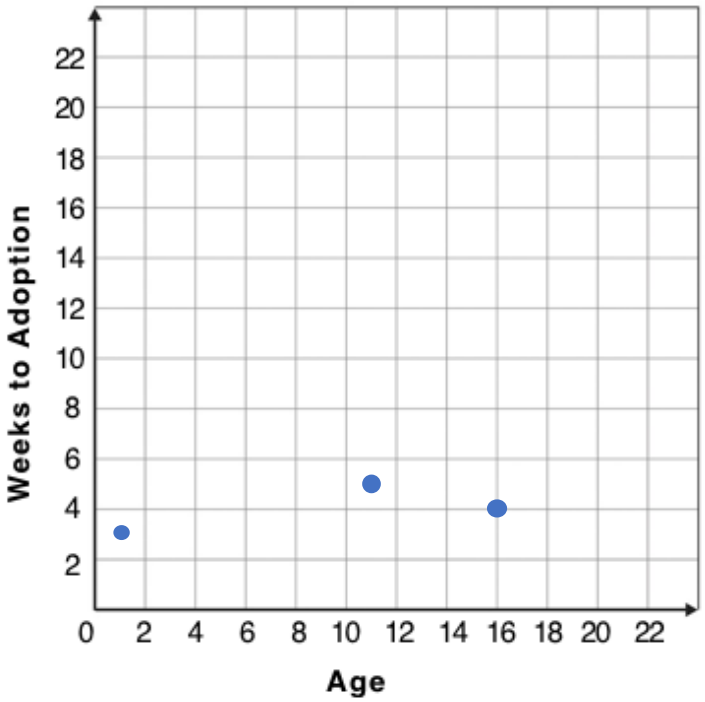
Do you agree? If so, why?

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? _____

Table Plan

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

Contract and Purpose

_____ **::** _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

_____ → _____

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) **:**

t = _____ *Define the table*

_____ *Are there more columns?*

_____ *Are there fewer rows?*

_____ *Are the rows ordered?*

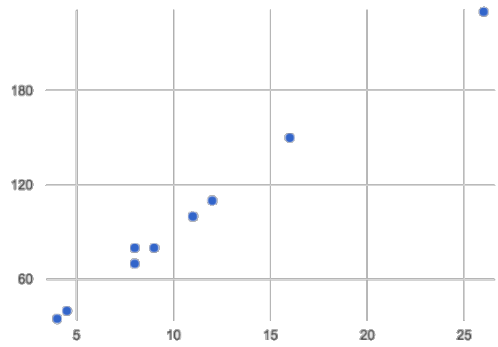
_____ *Produce the result*

end

Drawing Predictors

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

A

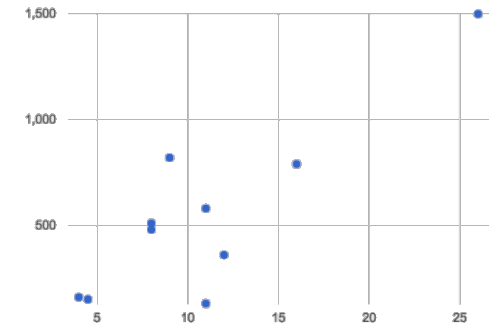


fat (g) v. calories-from-fat in common menu items

Direction: Positive Negative None

Strength: Strong Weak

B

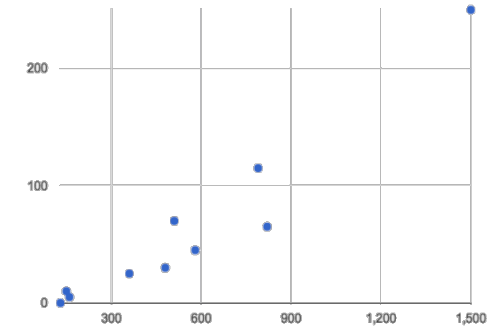


fat (g) v. sodium (g) in common menu items

Direction: Positive Negative None

Strength: Strong Weak

C

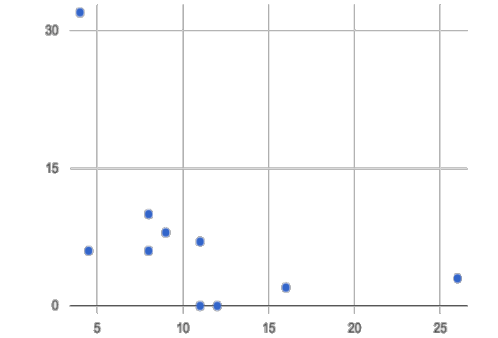


sodium (g) v. cholesterol (mg) in common menu items

Direction: Positive Negative None

Strength: Strong Weak

D



fat (g) v. sugar (g) in common menu items

Direction: Positive Negative None

Strength: Strong Weak

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 _____

_____. It would be stronger if I looked

at _____.
a subset or extension of my data

---

1) There may be a correlation between _____ and
column

_____. I think it is a _____, _____
column strong / weak positive / negative

correlation, because _____

_____. It would be stronger if I looked

at _____.
a subset or extension of my data

---

1) There may be a correlation between _____ and
column

_____. I think it is a _____, _____
column strong / weak positive / negative

correlation, because _____

_____. It would be stronger if I looked

at _____.
a subset or extension of my data

Unit 10

- 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 r^2. A **weak correlation** will have a small r^2.

- 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 r^2.

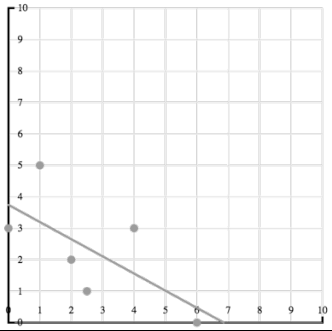
- **Correlation is not causation**! Correlation only suggests that two measures are *related*, but does not tell us if one *causes* the other. For example, hot days are *correlated* with people running their air conditioners, air conditioners do not *cause* hot days!

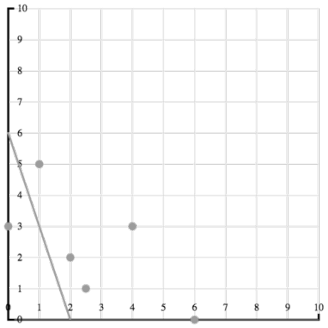
Grading Predictors

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)

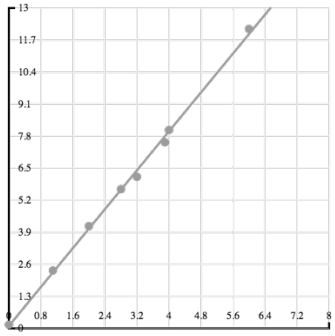
A

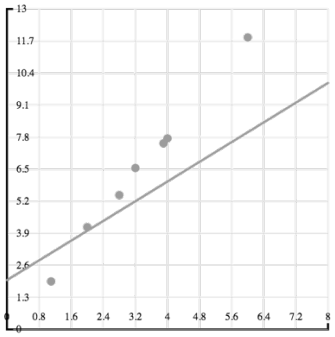




Strength of Correlation: _____

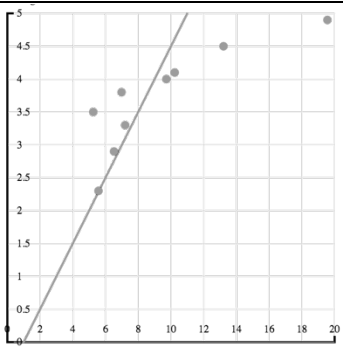
B

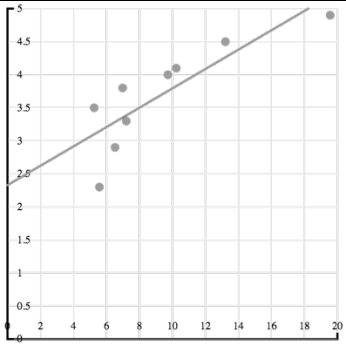




Strength of Correlation: _____

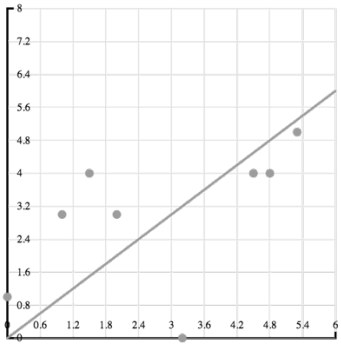
C

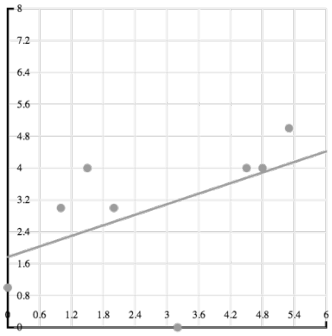




Strength of Correlation: _____

D





Strength of Correlation: _____

Findings in the `animals` Dataset

I performed a linear regression on dogs at the shelter, and
(dataset or subset)

found a weak (r^2=0.25), positive correlation between
(a strong/weak (r^2=…), positive/negative)

age of the dogs (in weeks) and number of weeks to be adopted. From this, I
([x-axis]) ([y-axis])

conclude that 25% of the variability in adoption time is explained
(r^2 % of the variation in [y-axis] is explained by [x-axis])

by the age of the dog.

I performed a linear regression on _____, and
(dataset or subset)

found _____ correlation between
(a strong/weak (r^2=…), positive/negative)

_____ and _____. From this, I
([x-axis]) ([y-axis])

conclude that _____
(r^2 % of the variation in [y-axis] is explained by [x-axis])

_____.

I performed a linear regression on _____, and
(dataset or subset)

found _____ correlation between
(a strong/weak (r^2=…), positive/negative)

_____ and _____. From this, I
([x-axis]) ([y-axis])

conclude that _____
(r^2 % of the variation in [y-axis] is explained by [x-axis])

_____.

Correlations in My Dataset

I performed a linear regression on _____, and
dataset or subset

found _____ correlation between
a strong/weak (r^2=…), positive/negative

_____ and _____. From this, I
[x-axis] [y-axis]

conclude that _____
r^2 % of the variation in [y-axis] is explained by [x-axis]

_____.

I performed a linear regression on _____, and
dataset or subset

found _____ correlation between
a strong/weak (r^2=…), positive/negative

_____ and _____. From this, I
[x-axis] [y-axis]

conclude that _____
r^2 % of the variation in [y-axis] is explained by [x-axis]

_____.

I performed a linear regression on _____, and
dataset or subset

found _____ correlation between
a strong/weak (r^2=…), positive/negative

_____ and _____. From this, I
[x-axis] [y-axis]

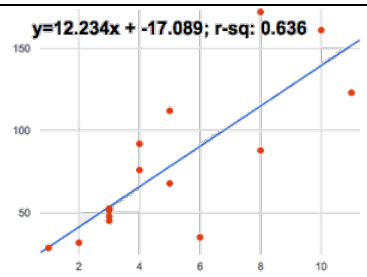
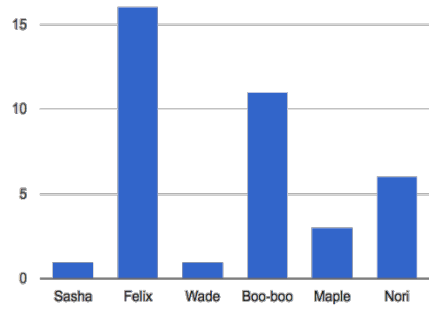
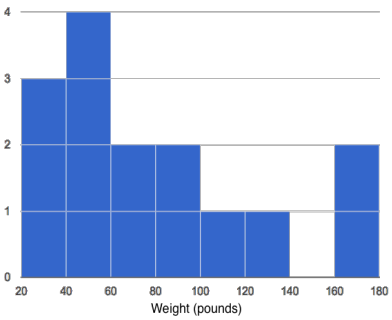
conclude that _____
r^2 % of the variation in [y-axis] is explained by [x-axis]

_____.

Unit 11

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'."*	
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		*"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	 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		*"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

```
_________________::_________________________ → _________________
        name                  domain                    range
# ____________________________________________________________
examples:

    ____________(________) is ______________________________

    ____________(________) is ______________________________
end

fun ____________ (_______) :  ______________________________

end
```

```
_________________::_________________________ → _________________
        name                  domain                    range
# ____________________________________________________________
examples:

    ____________(________) is ______________________________

    ____________(________) is ______________________________
end

fun ____________ (_______) :  ______________________________

end
```

Design Recipes

```
_________________::_________________________ → _________________
        name                  domain                    range
# ____________________________________________________________
examples:

    ____________(________) is ____________________________

    ____________(________) is ____________________________
end

fun ____________ (_______) :  ____________________________

end
```

```
_________________::_________________________ → _________________
        name                  domain                    range
# ____________________________________________________________
examples:

    ____________(________) is ____________________________

    ____________(________) is ____________________________
end

fun ____________ (_______) :  ____________________________

end
```

Design Recipes

```
_________________::_________________________ → _________________
        name                  domain                    range
# ____________________________________________________________
examples:

    ____________(________) is ______________________________

    ____________(________) is ______________________________
end

fun ____________ (_______) :  ______________________________

end
```

```
_________________::_________________________ → _________________
        name                  domain                    range
# ____________________________________________________________
examples:

    ____________(________) is ______________________________

    ____________(________) is ______________________________
end

fun ____________ (_______) :  ______________________________

end
```

Table Plan

Contract and Purpose

_____ : : _____ → _____

_____

_____

Examples

Make a Start Table and a result based on that table.

_____ → _____

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) :

t = _____

_____

_____

_____

_____

end

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

Table Plan

Contract and Purpose

_____ **::** _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

_____ → _____

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) **:**

t = _____

_____

_____

_____

_____

end

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

Table Plan

Contract and Purpose

_____ : : _____ → _____

_____

Examples

Make a Start Table and a result based on that table.

_____ → _____

Define the function

Use the relevant methods (circle your helper functions!), then produce a result with the new table.

fun _____ (_____) :

t = _____

_____

_____

_____

_____

end

Define the table

Are there more columns?

Are there fewer rows?

Are the rows ordered?

Produce the result

Contracts

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

Contracts

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