



BOOTSTRAP

www.bootstrapworld.org

Data Science



Workbook v0.9b

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

The question I chose is...

I think the answer is...

The data I would use to support my answer is...

Expressions, Values, and Errors

For each expression, if it produces an error when evaluated, write what kind of error occurs:

- For division by zero errors, write "division by 0".
- For errors where the operator is given the wrong type, write "wrong type".
- Otherwise, write what the expression evaluates to.

Expression	Value, or Error?
$8 - 5.3$	
$2 / 0$	
<code>"Three" * 2</code>	
$(3 + 5) * 3$	
$1.5 * "6"$	
$(2 / (3 - (2 + 1)))$	

Definitions

For each of the definitions below, create values that match the type shown in the comment (after the #).

```
age    = _____ # age :: Number
big    = _____ # bigNum :: Number
small  = _____ # smallNum :: Number
name   = _____ # name :: String
city   = _____ # city :: String
state  = _____ # state :: String
evens  = [list:_____ # evens :: List<Number>
odds   = _____ # odds :: List<Number>
colors = _____ # colors :: List<String>
```

Unit 2

“What is the relationship between calories and sugar?”

I hypothesize...

I found...

Animals

Animal	Number-of-legs
"Human"	2
"Ant"	6
"Spider"	8
"Bear"	4
"Snake"	0

1. How many rows does this table have? _____
2. How many columns does this table have? _____
3. What are the names of the columns? _____
4. For the row with value "Human" in the **Animal** column, what is the value in the **Number-of-legs** column? _____
5. Circle the header row of this table

Presidents and Nutrition

Answer the following questions about the `presidents` and `nutrition` tables, using your Unit-2 Pyret program:

1. How many columns does the `presidents` table have? _____
2. What are the names of the columns? _____
3. How many rows does the `presidents` table have? _____
4. Is the `party` column quantitative or categorical? _____
5. Is the data in the `home-state` column categorical? _____
6. If so, how many categories are there? _____
7. What is the home state of Millard Fillmore? _____
8. Who was the first president from the Federalist party? _____
9. How many columns does the `nutrition` table have? _____
10. How many rows does the `nutrition` table have? _____
11. How many grams of cholesterol does the Hamburger have? _____
12. Which food has the largest serving size? _____
13. Is the data in the `calories` column quantitative? If so, why?

Unit 3

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Reading Charts

1. Which menu item has the most sodium? _____
2. Which menu item has the least sodium? _____
3. Do french fries have more sodium than hamburgers? _____
4. Which country has the largest GDP? _____
5. What percent of the total world GDP is from China? _____

Frequency Bar Chart

First	Last	Eye-Color
"John"	"Doe"	"Green"
"Jane"	"Smith"	"Brown"
"Javon"	"Jackson"	"Brown"
"Angela"	"Enriquez"	"Hazel"
"Jack"	"Thompson"	"Blue"
"Dominique"	"Rodriguez"	"Hazel"
"Sammy"	"Carter"	"Blue"
"Andrea"	"Garcia"	"Brown"

1. How many students have Brown eyes? _____

2. How many students have Green eyes? _____

3. How many students have Hazel eyes? _____

4. How many students have Blue eyes? _____

5. Above the "Blue" label on this bar chart, add a bar with height that corresponds to the number of students with Blue eyes.

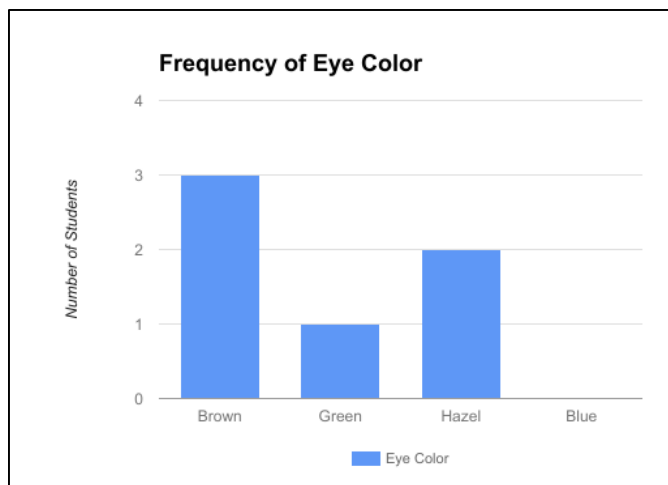
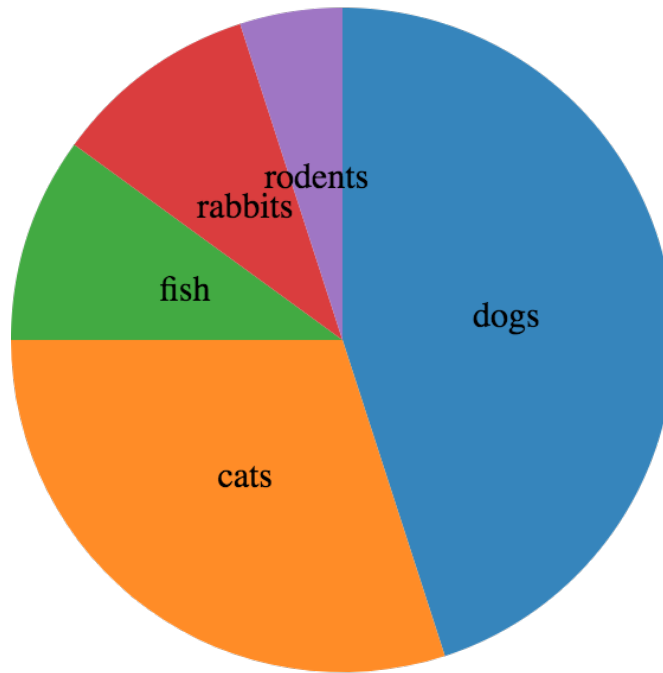
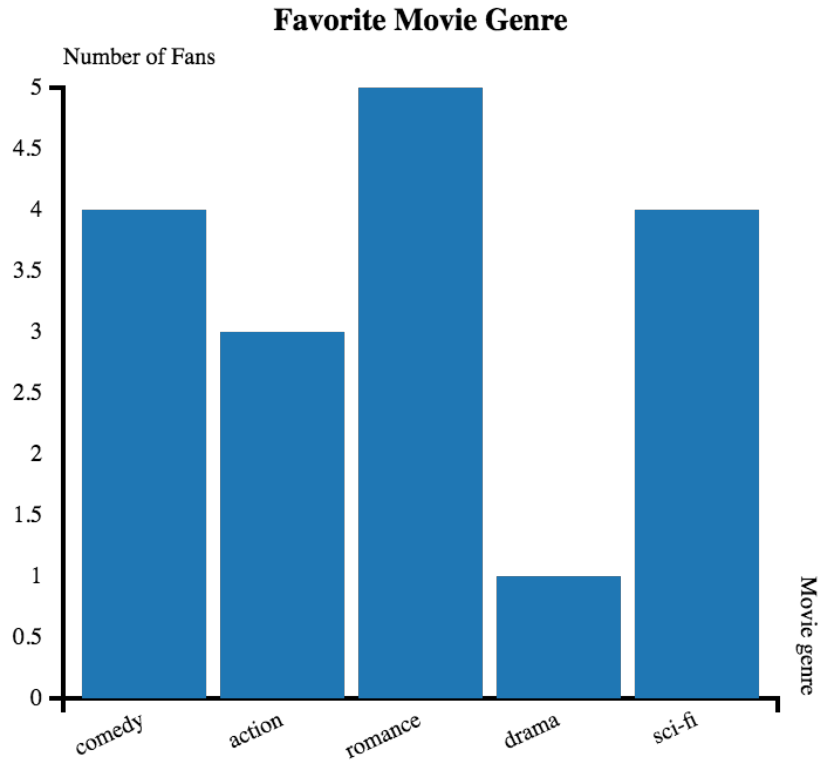


Chart Practice

Pet Ownership

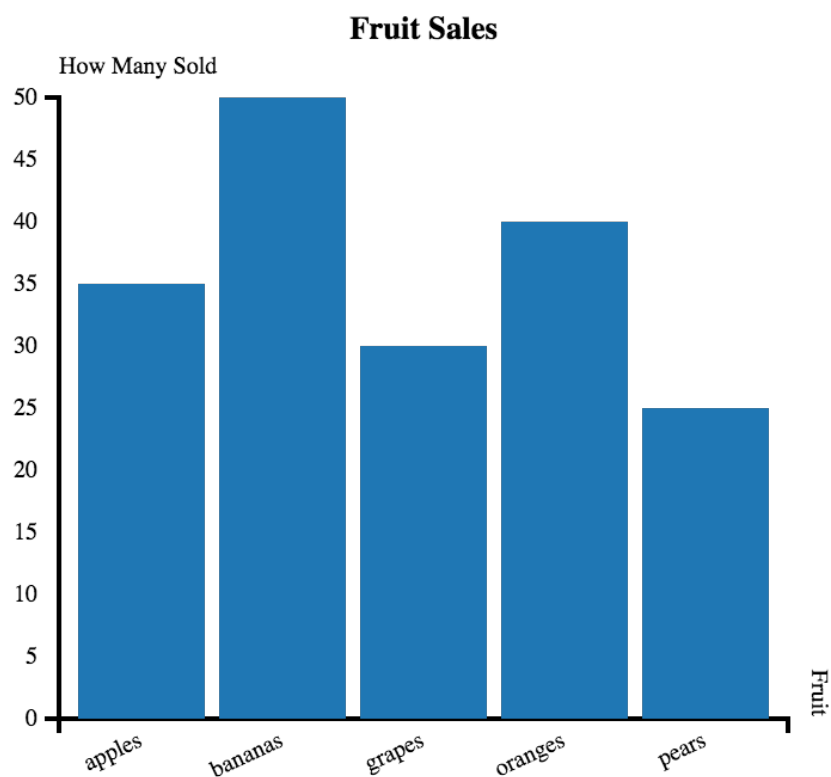


1. Is this a pie chart, or a bar chart? _____
2. Which pet is the most popular? _____
3. Which pet is the least popular? _____
4. Which are more popular, fish or rodents? _____



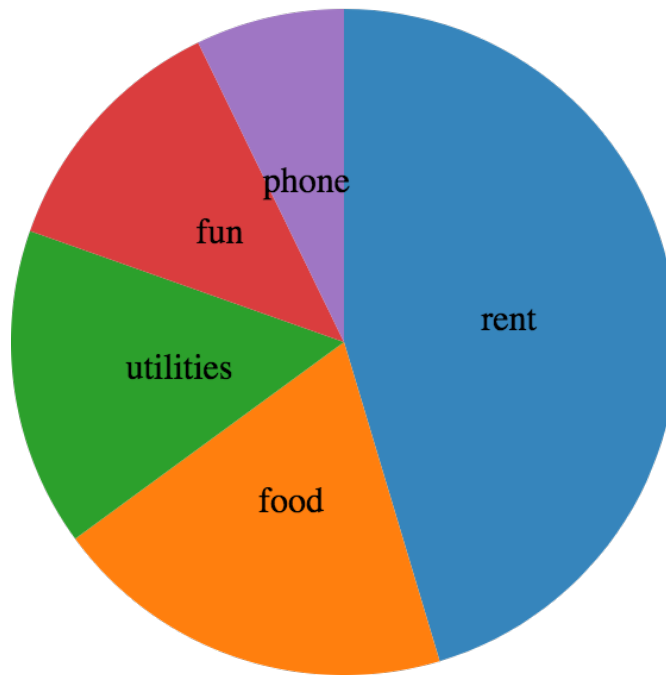
1. Is this a bar chart or a pie chart? _____
2. What movie genre is most popular? _____
3. What are the labels of this chart? _____
4. What are the values of this chart? _____
5. Is this a frequency bar chart? _____

More Chart Practice



1. Are apples more popular than grapes? _____
2. How many categories of fruit are there? _____
3. How many pears were sold? _____
4. What fruit is least popular? _____

Monthly Budget



1. Which expense needs the least amount of money? _____
2. Which expense takes up almost half of the budget? _____
3. Suppose a person has a \$2000 monthly budget, and they spend 15% on food. How many dollars do they spend on food in a single month? _____

Unit 4

“According to the US Census Bureau, the average American household earned more than \$45,000 in 2003 - more than 3x the poverty line that year. Does that mean only a small percentage of Americans were in poverty that year?”

I hypothesize...

I found...

Mean, Median, Mode Practice

Using pencil & paper, calculate the 3 numbers that measure the center of each list. If a list contains more than one mode, write the number with the smallest value.

These lists are bound to variables a, b, c, d, e in the Unit 4 template file, so you can check your answers with Pyret.

List	Mean	Median	Mode-List
a = [list: 1, 1, 4]			[list:]
b = [list: 3, 4, 5]			[list:]
c = [list: 3, 3, 4, 6]			[list:]
d = [list: -1,0.5,2,0.5,2,6]			[list:]
e = [list: 2, 11, 7, 4]			[list:]

Measuring Center in Pyret

1. What is the mode of the `calories-list`? _____
2. What is the mean amount of `sodium` for menu items? _____
3. What is the median GDP for all the countries in `countries`? _____
4. What is the median of `life-expectancy-list`? _____

Imagine the following code is in your definitions window:

```
mystery-list = [list: 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

5. What is the median of this `mystery-list`? _____

Now imagine these lists (which contain the same elements as `mystery-list`) are in your definitions window:

```
mystery1 = [list: 1, 4, 7]  
mystery2 = [list: 2, 3, 8]  
mystery3 = [list: 5, 6, 9]
```

6. What is the median of `mystery1`? _____
7. What is the median of `mystery2`? _____
8. What is the median of `mystery3`? _____
9. What is the median of a list containing these 3 medians? _____
10. Is this different from the median of `mystery-list`? _____

Unit 5

Roll two dice, and guess the sum of the roll. Guess right and you win \$1. Guess wrong and you lose \$1.

"Is this game worth playing?"

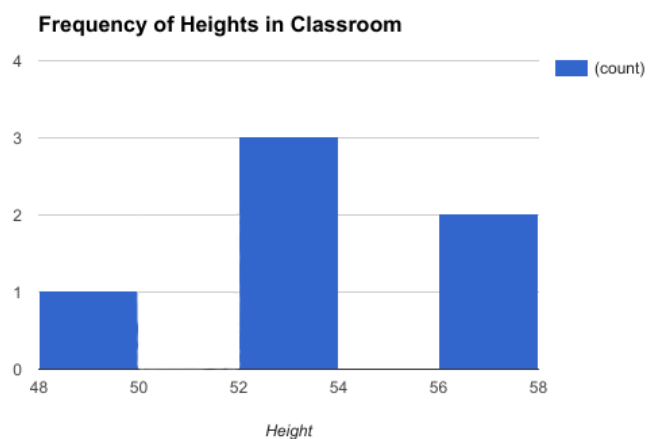
I hypothesize...

I found...

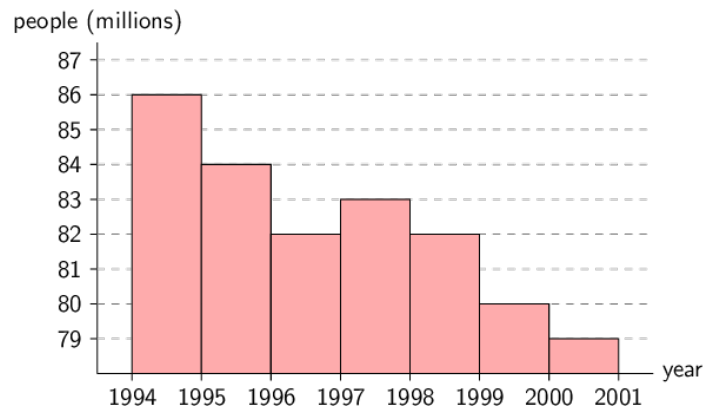
Histograms

First	Last	Height
"John"	"Doe"	52.0
"Jane"	"Smith"	49.1
"Javon"	"Jackson"	57.7
"Angela"	"Enriquez"	52.5
"Jack"	"Thompson"	53.0
"Dominique"	"Rodriguez"	51.1
"Sammy"	"Carter"	56.2
"Andrea"	"Garcia"	50.8

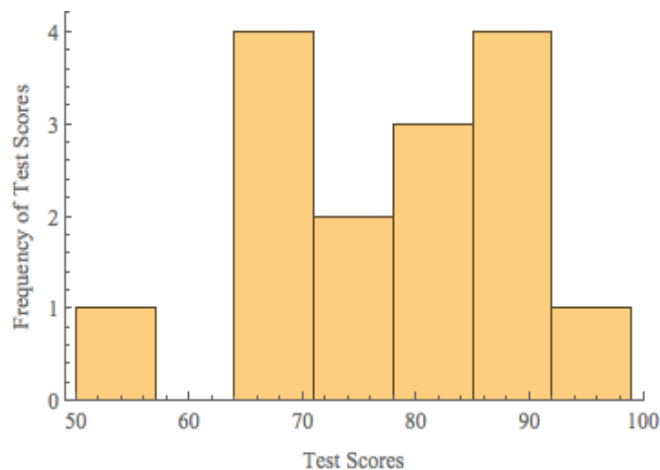
- How many students are 48in or above, *and* less than 50in tall? _____
- How many students are 50in or above, *and* less than 52in tall? _____
- How many students are 52in or above, *and* less than 54in tall? _____
- How many students are 54in or above, *and* less than 56in tall? _____
- How many students are 56in or above, *and* less than 58in tall? _____
- Add a bar to this histogram for students who are 50in or above, *and* less than 52in tall.



Histogram Practice



1. How many people were born between 1996 and 1997? _____
2. On what year were the most number of people born? _____
3. How many bins does this histogram have? _____
4. Were more people born in 1994 or 1995? _____



1. How many bins does this histogram have? _____
2. What is (are) the bins with the highest frequency of scores? _____
3. How many students scored between 85 and 92? _____

Unit 6

"The more you pay at a restaurant, the better it is."

Do you agree?

I hypothesize...

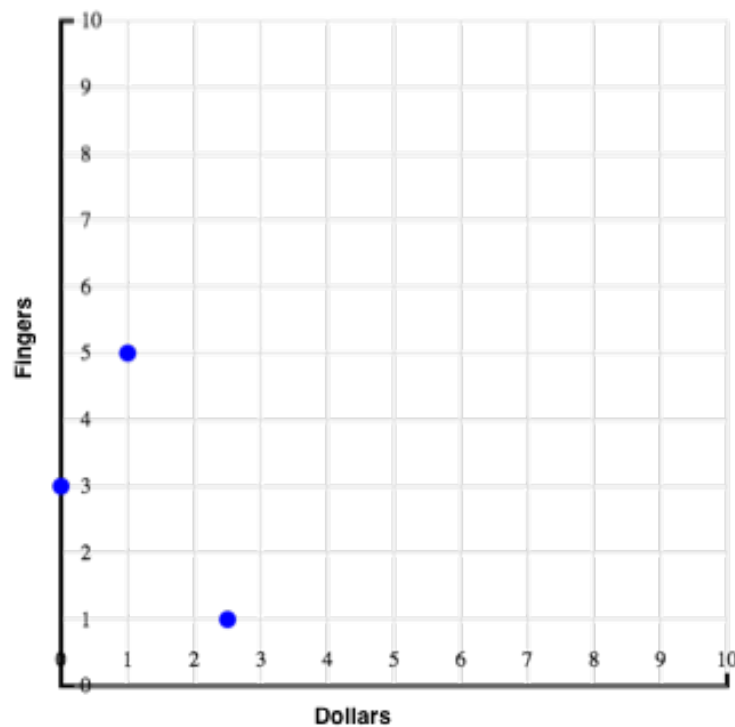


I found...

Creating a Scatter Plot

For each row in the following table, add a dot to the scatter plot. The first 3 rows have been completed for you. Use the values from the left column along the horizontal axis, and values from the right column along the vertical axis.

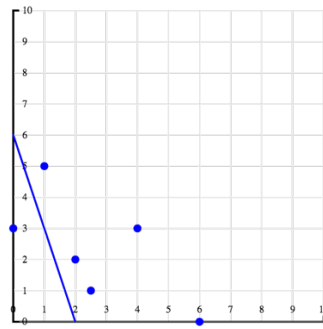
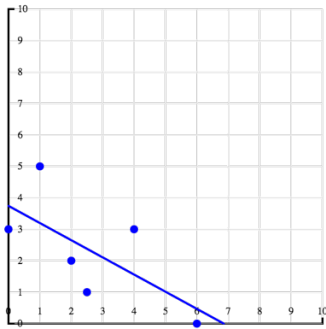
Dollars	Fingers
0	3
1	5
2.5	1
2	2
6	0
4	3



Grading Predictor Functions

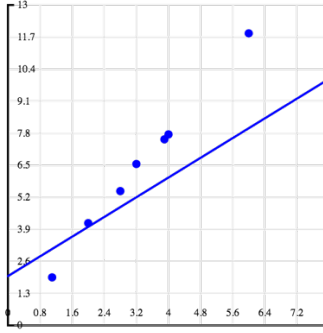
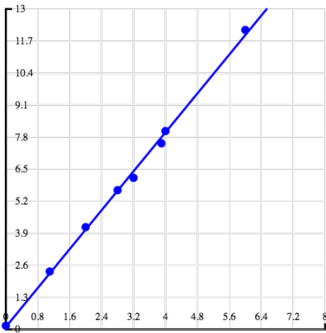
Below are the scatterplots for 4 data sets, with two different predictors shown for each set. For each data set, **circle the plot with the predictor function that fits better**, and **give it a grade between 0 (worst possible fit) and 1 (best possible fit)**.

1



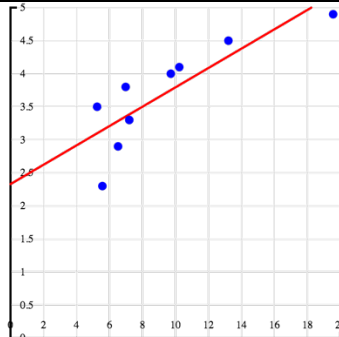
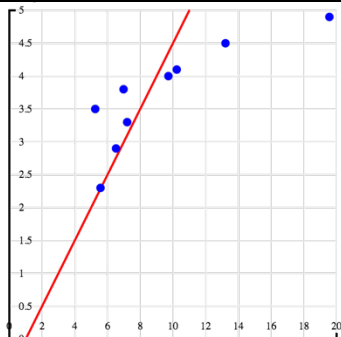
Grade for best predictor:

2



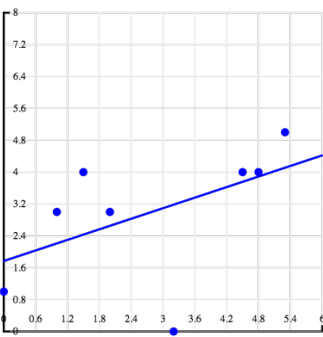
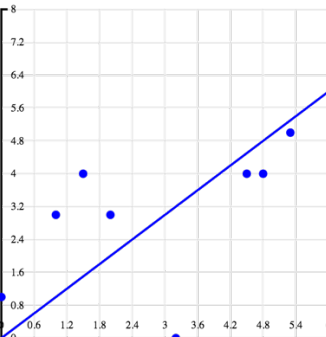
Grade for best predictor:

3



Grade for best predictor:

4



Grade for best predictor:

Checking for Understanding

1. In your own words, explain what a **predictor function** is.

2. In your own words, explain what the **r-squared** value of a predictor is.

Unit 7

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Practice with Select

Below is a table bound to the variable name `animals`.

name	legs	eyes	lifespan
"Human"	2	2	71
"Garden Ant"	6	2	8
"Spider"	8	8	2.5
"Bear"	4	2	10

1. Draw the table produced by this code (don't forget the header row!):

```
animals.select-columns([list: "lifespan", "name"])
```


2. What code will produce the table shown here?

eyes
2
2
8
2

3. Challenge: Draw `table2`, produced by this code:

```
table1 = animals.select-columns([list: "name", "legs"])
```

```
table2 = table1.select-columns([list: "legs"])
```

table2

Table Plan: What do we need?

We can use tables to do all sorts of things – but we need a plan. Each of the following questions involves some subset of the `animals` table. Read each one carefully, then write a table query that will *remove unnecessary columns* – keeping only those we need – and binds the new table to a variable you choose.

animals

name	legs	eyes	lifespan
"Human"	2	2	71
"Garden Ant"	6	2	8
"Spider"	8	8	2.5
"Bear"	4	2	10

1. We want to make a histogram showing the distribution of `legs`

Do I want to select only certain columns?

_____ =

2. We want to make a scatterplot of the relationship between `legs` and `eyes`.

Do I want to select only certain columns?

_____ =

3. We want to search for a predictor function linking `eyes` and `lifespan`

Do I want to select only certain columns?

_____ =

Table Plan: Is there an order?

We can use tables to do all sorts of things – but we need a plan. Each of the following questions involves the `animals` table. Read each one carefully, then write a table query that will *orders the rows of the table* – in the correct order – and binds the new table to a variable you choose.

animals

name	legs	eyes	lifespan
"Human"	2	2	71
"Garden Ant"	6	2	8
"Spider"	8	8	2.5
"Bear"	4	2	10

1. We want a table that has the shortest-lived animal first and longest-lived last.

Do I want to order by a specific column?

_____ =

2. We want to extract a list of legs, from most-to-least.

Do I want to order by a specific column?

_____ =

3. We want an alphabetized list of animal names.

Do I want to order by a specific column?

_____ =

Table Plan: Total and Domestic

We'd like to sort our movies in ascending order of `total`, and then show only the `title`, `total`, and `domestic`.

(The table on the left is a **sample table**, containing a few rows from the full table. This is a small sample we can start from. The **sample table** on the right is where we need to end up. Your job is to write the queries that get us there.)

movies

Movie Title	Studio	Total	Domestic	Overseas	Year
Interstellar	Par.	675.1	188	487.1	2014
The Sixth Sense	BV	672.8	293.5	379.3	1999
Man of Steel	WB	668	291	377	2013
Kung Fu Panda 2	P/DW	665.7	165.2	500.4	2011
Ice Age: The Meltdown	Fox	660.9	195.3	465.6	2006



total-and-domestic

Movie Title	Total	Domestic
Ice Age: The Meltdown	660.9	195.3
Kung Fu Panda 2	665.7	165.2
Man of Steel	668	291
The Sixth Sense	672.8	293.5
Interstellar	675.1	188

Do I want to order by a specific column?

_____ =

Do I want to select only certain columns?

_____ =

Table Plan: Title and Year

We'd like to sort our movies in descending order of `year`, and then show only the `title` and `year`.

(The table on the left is a **sample table**, containing a few rows from the full table. This is a small sample we can start from. The **sample table** on the right is where we need to end up. Your job is to write the queries that get us there.)

movies

Movie Title	Studio	Total Gross	Domestic	Overseas	Year
Interstellar	Par.	675.1	188	487.1	2014
The Sixth Sense	BV	672.8	293.5	379.3	1999
Man of Steel	WB	668	291	377	2013
Kung Fu Panda 2	P/DW	665.7	165.2	500.4	2011
Ice Age: The Meltdown	Fox	660.9	195.3	465.6	2006



title-and-year

Title	Year
Interstellar	2014
Man of Steel	2013
Kung Fu Panda 2	2011
Ice Age: The Meltdown	2006
The Sixth Sense	1999

Do I want to order by a specific column?

_____ =

Do I want to select only certain columns?

_____ =

Unit 8

"How much of Asia's GDP does China generate?"

I hypothesize...

I found...

Booleans and Comparison

Suppose your program has the following definitions:

```
legs = 2
eyes = 2
class = "Mammal"
continent = "North America"
```

What will each of the following expressions evaluate to?

Expression	Value
<code>legs <= 4</code>	
<code>eyes == 2</code>	
<code>legs <> 4</code>	
<code>eyes <> 5 - 3</code>	
<code>legs == eyes</code>	

When you finish the first table try these challenge questions:

Expression	Value
<code>class == "Mammal"</code>	
<code>class == "Invertebrate"</code>	
<code>class <> "mammal"</code>	
<code>continent == "Asia"</code>	

Table Plan: Recent Title and Year

Show the title and year for movies released after 2010, in descending order of total gross.

movies

Movie Title	Studio	Total	Domestic	Overseas	Year
Interstellar	Par.	675.1	188	487.1	2014
The Sixth Sense	BV	672.8	293.5	379.3	1999
Man of Steel	WB	668	291	377	2013
Kung Fu Panda 2	P/DW	665.7	165.2	500.4	2011
Ice Age: The Meltdown	Fox	660.9	195.3	465.6	2006

solution4

Title	Year
Interstellar	2014
Man of Steel	2013
Kung Fu Panda 2	2011



Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Table Plan: Title and Overseas

Starting with the table below, produce a table of `Titles` and `Overseas` profits, for all movies made before 2010, in ascending order of `Total Gross`.

Note: Start by filling in what the *solution* table should look like!

movies-start

Movie Title	Studio	Total Gross	Domestic	Overseas	Year
Interstellar	Par.	675.1	188	487.1	2014
The Sixth Sense	BV	672.8	293.5	379.3	1999
Man of Steel	WB	668	291	377	2013
Kung Fu Panda 2	P/DW	665.7	165.2	500.4	2011
Ice Age: The Meltdown	Fox	660.9	195.3	465.6	2006



solution5

Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Bad Starter Tables!

For each of the questions below, find out what's wrong with the provided starter table. Write your answer in space below each table.

1. "Make a table of all the presidents, sorted alphabetically by home-state"

nth	name	home-state	yr-started	yr-ended	Party
7	Andrew Jackson	Tennessee	1829	1837	Democratic

2. "Make a table showing only Democratic Presidents"

nth	name	home-state	yr-started	yr-ended	party
7	Andrew Jackson	Tennessee	1829	1837	Democratic
35	John F. Kennedy	Massachusetts	1961	1963	Democratic
11	James K. Polk	Tennessee	1845	1849	Democratic
44	Barack Obama	Illinois	2009	2017	Democratic

3. "Make a table showing the presidents sorted in ascending order of year-started"

nth	name	home-state	yr-started	yr-ended	party
22	Grover Cleveland	New York	1885	1889	Democratic
24	Grover Cleveland	New York	1893	1897	Democratic

4. "Make a table showing all presidents from New York."

nth	name	home-state	yr-started	yr-ended	party
45	Donald Trump	New York	2017	2021	Republican
32	Franklin D. Roosevelt	New York	1933	1945	Democratic
21	Chester A. Arthur	New York	1881	1885	Republican
26	Theodore Roosevelt	New York	1901	1909	Republican

Table Plan: Asian GDPs

Define a table showing the names and GDPs of all countries in Asia, starting with the `countries` table.

Start out by creating a realistic “starter table”, using a sample of rows from the `countries` table, then a desired “end table” showing only the rows and columns you want, in the order you want them.

countries

asian-GDPs



Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Unit 9

“Is GDP-per-capita positively correlated with life expectancy?”

I hypothesize...

I found...

Extending Tables

Below is a table called `games`, which contains the number of points scored by different NBA players in their first 3 games of a season. Complete the new table on the right by filling in the value of the **total** column (just add the **game1**, **game2**, **game3** columns together).

games

player	game1	game2	game3
"Lebron James"	30	28	36
"Steph Curry"	26	32	29
"Kyrie Irving"	21	24	27
"John Wall"	27	30	25
"Isaiah Thomas"	25	22	24



games-with-total

player	game1	game2	game3	total
"Lebron James"	30	28	36	
"Steph Curry"	26	32	29	
"Kyrie Irving"	21	24	27	
"John Wall"	27	30	25	
"Isaiah Thomas"	25	22	24	

1. Which player has scored the most points so far? _____

Below is a table named `socks`, containing the prices of *packs* of socks at several different stores. Each store sells different size packs, for different prices. Complete the new table on the right by filling in the value of the **price-per-sock** column.

socks

name	price	socks
"Super Store"	2.50	4
"Clothes Galore"	5.40	4
"Bargain Mart"	4.50	6
"Fashion Statement"	15.00	12
"Sock Emporium"	7.00	10



socks-with-proce

name	price	socks	price-per-sock
"Super Store"	2.50	4	
"Clothes Galore"	5.40	4	
"Bargain Mart"	4.50	6	
"Fashion Statement"	15.00	12	
"Sock Emporium"	7.00	10	

2. Which store has the best deal on socks? _____

Table Plan: Body Building

Your aunt is a bodybuilder, and wants to eat only foods that have at least .12 grams of protein per serving. Starting with nutrition, build a table showing only the name, calories and protein-per-gram for menu items that fit this criterion.

(Suggestion: draw a start and end sample table on a sheet of scrap paper!)

Do I need to *build* a column?

_____ =

Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Table Plan: Term Length

For how many years was each Democratic president in office? We'd like to make a histogram showing how many democratic presidents served between 0 - 4 years, or 4 - 8 years. How do we make the necessary table?

Do I need to *build* a column?

_____ =

Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Table Plan: GDP v. Population

The United Nations wants us to investigate whether per-capita-gdp or population size has a larger influence on median life expectancy in Africa.

(Suggestion: draw a start and end sample table on a sheet of scrap paper!)

Do I need to *build* a column?

_____ =

Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Countries Table Plan Practice

Make a histogram of per-capita GDP for countries with universal health care.
Do most of these countries have a per-capita GDP that is higher than the average per-capita GDP of all countries?

Do I need to *build* a column?

_____ =

Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Table Plan

Do I need to *build* a column?

_____ =

Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Table Plan

Do I need to *build* a column?

_____ =

Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Table Plan

Do I need to *build* a column?

_____ =

Do I want to *filter* just certain rows?

_____ =

Do I want to *order* by a specific column?

_____ =

Do I want to *select* only certain columns?

_____ =

Contracts

Name	Domain	Range
triangle	:: (side :: <i>Number</i> , mode :: <i>String</i> , color :: <i>String</i>)	→ <i>Image</i>
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→
	::	→

Contracts

[illegible]

