BOOTSTRAP

www.bootstrapworld.org

Data Science

Pilot Workbook – Summer 2017

BOOTSTRAP

www.bootstrapworld.org

Workbook v0.9b

Brought to you by the Bootstrap team:

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

Unit 1

(and room for notes!)

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`	
`"Three" * 2`	
`(3 + 5) * 3`	
`1.5 * "6"`	
`(2 / (3 - (2 + 1)))`	

Identifiers and Expressions

Imagine the program below has been written in your definitions window:

```
x = (3 * 2) - 2
y = x * 1.5
```

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 a variable hasn't been defined, write "unbound id"
- Otherwise, write what the expression evaluates to.

Expression	Value, or Error?
`y`	
`x - 3`	
`(y - 1) * z`	
`(x + y) / 2`	
`x + y`	

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

"The average US Household makes more than $45,000/yr[1]. So why are so many people living in poverty?"

I hypothesize…

I found…

[1] https://web.archive.org/web/20060903121944/http://www.census.gov/hhes/income/histinc/h13.html

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 3 template file, so you can check your answers with Pyret.

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

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 4

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.

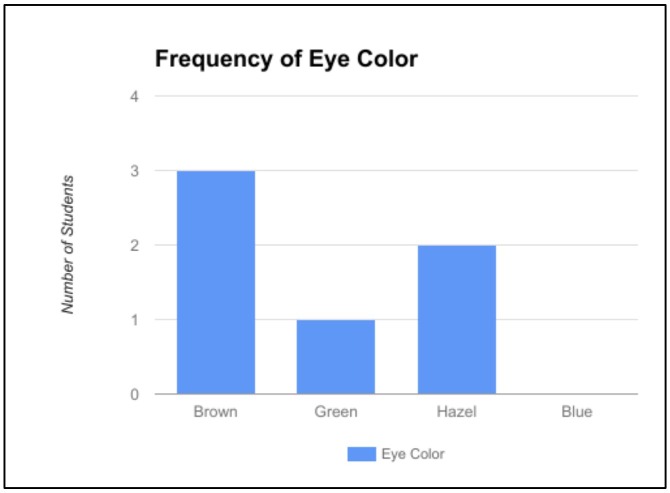
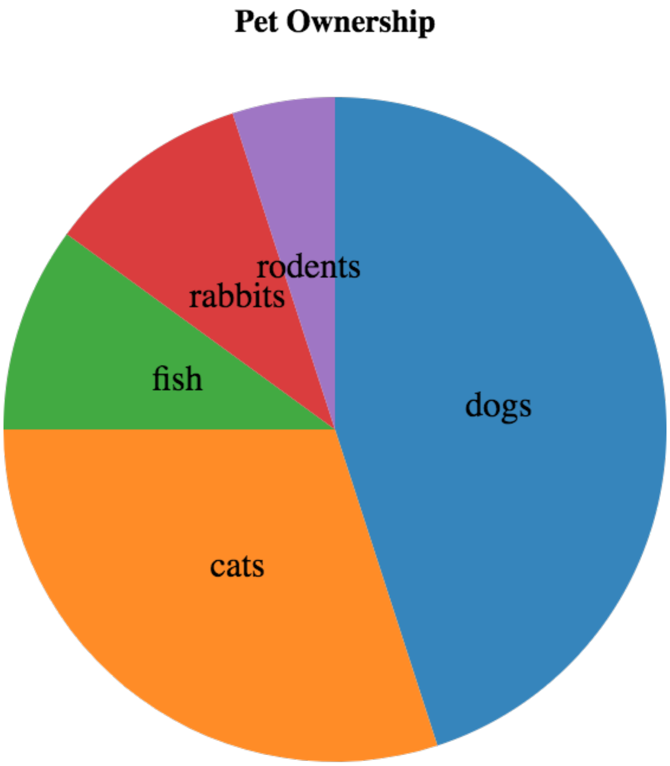


Chart Practice

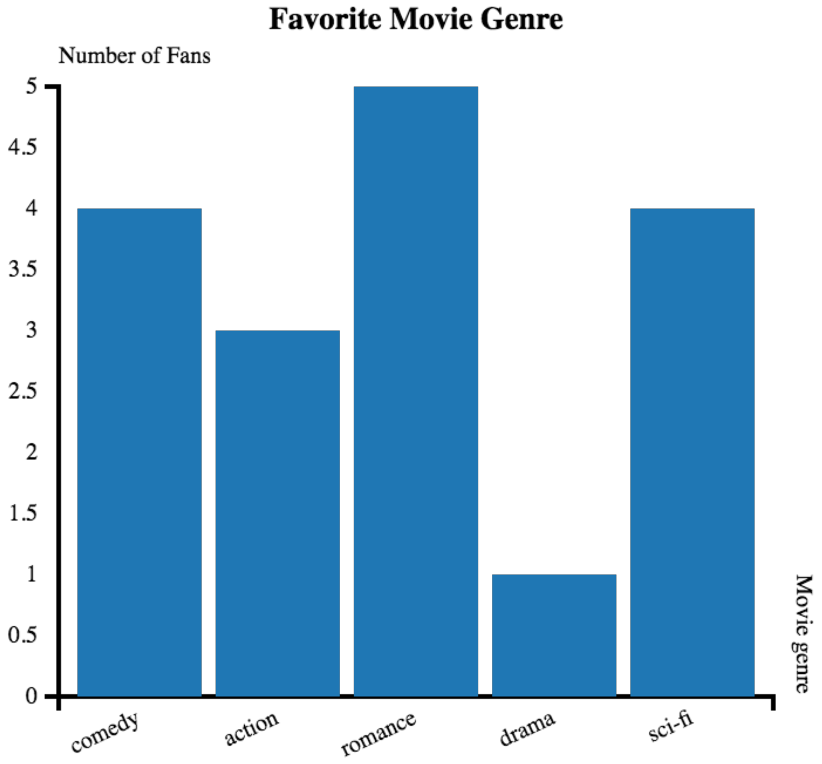


1. Is this a pie chart, or a bar chart? ________________________

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? _________________________

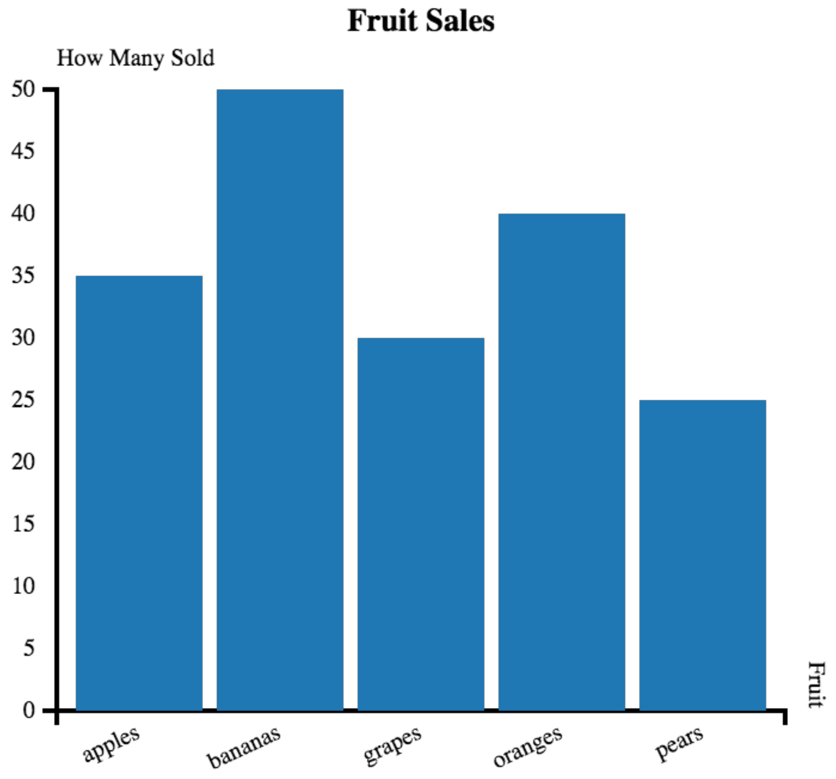
2. What genre is most popular? _________________________

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



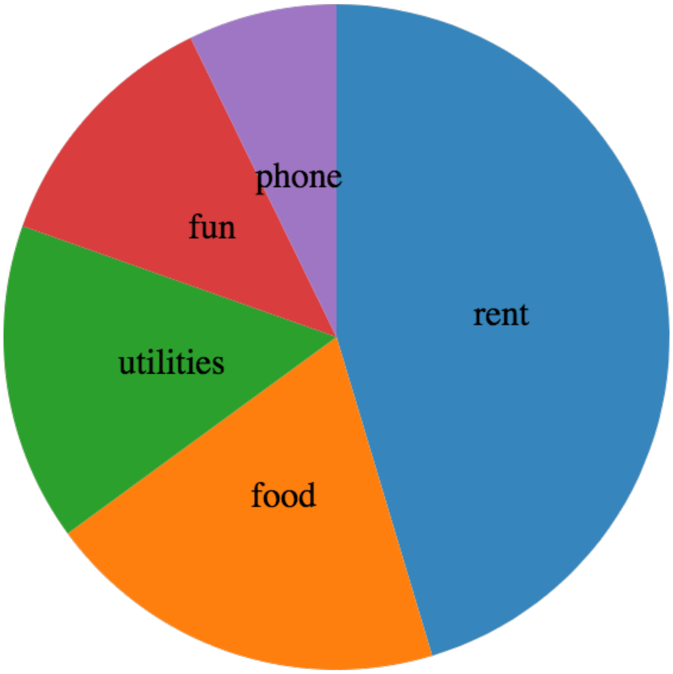
1. Are apples more popular than grapes? ________________

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 is spent on food in a single month? _______________

Unit 5

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

"What are your chances of winning?"

I hypothesize…

I found…

Introducing 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

1. How many students are between 48 and 50 inches tall? ______________

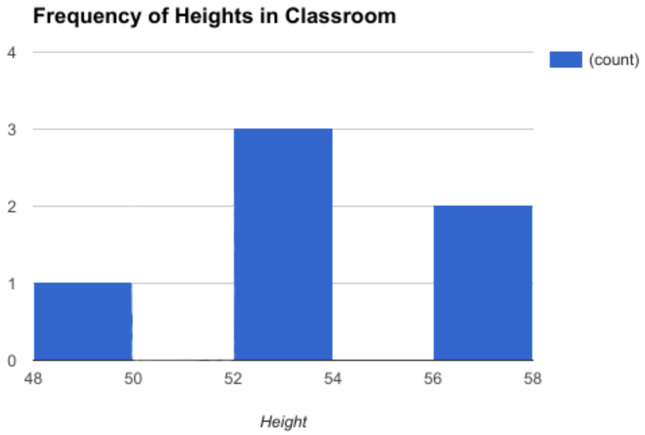
2. How many students are between 50 and 52 inches tall? ______________

3. How many students are between 52 and 54 inches tall? ______________

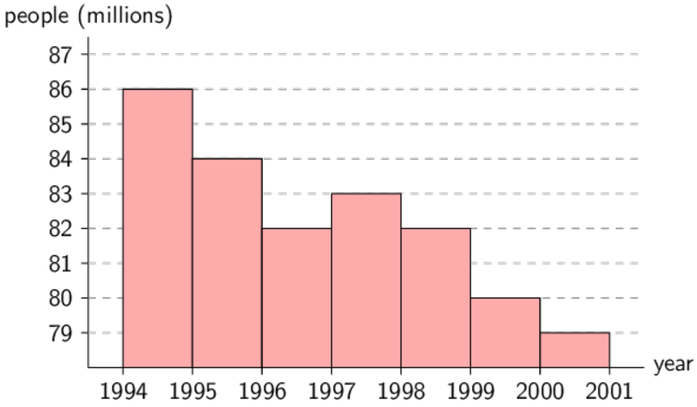
4. How many students are between 54 and 56 inches tall? ______________

5. How many students are between 56 and 58 inches tall? ______________

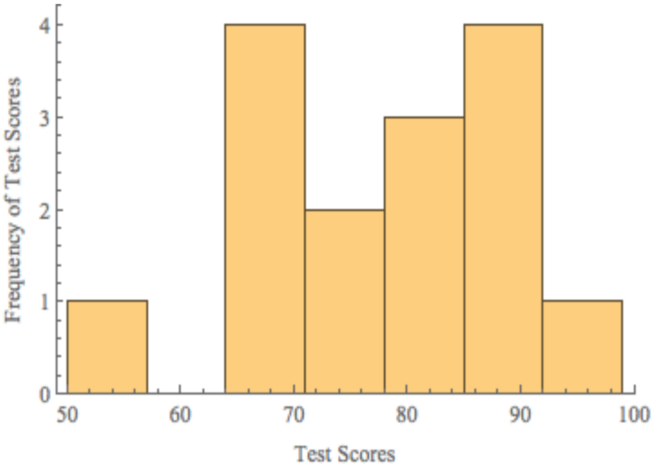
6. Add a bar to this histogram for students who are between 50 and 52 inches tall.



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

"Are more expensive restaurants generally better than cheaper ones?"

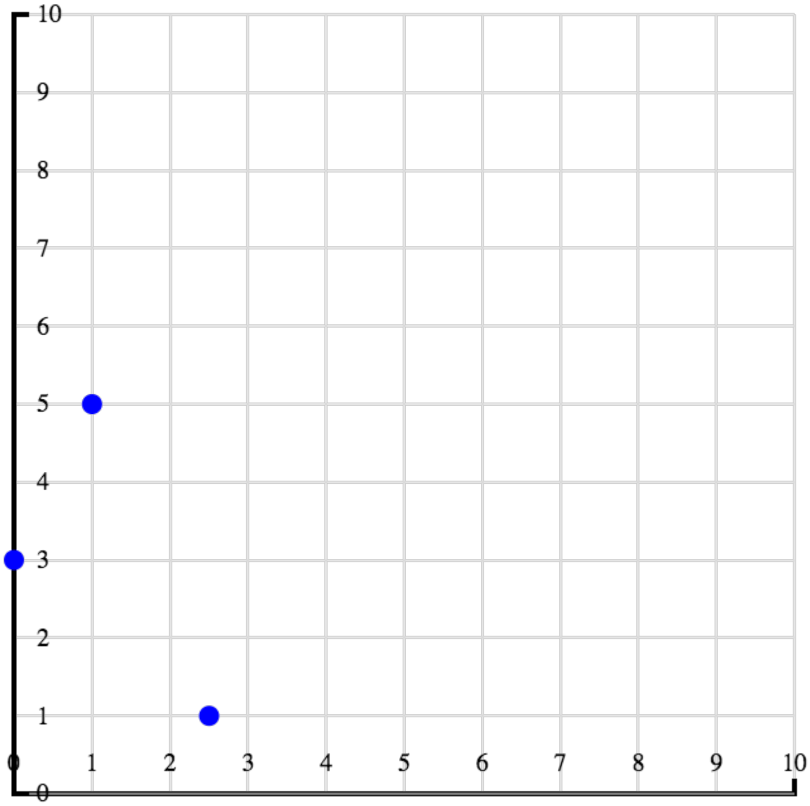
I hypothesize…

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.

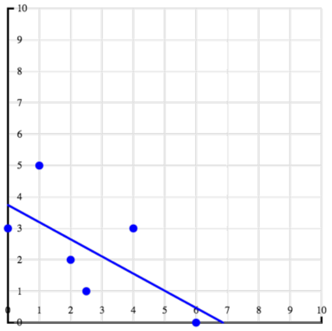
0	3
1	5
2.5	1
2	2
6	0
4	3

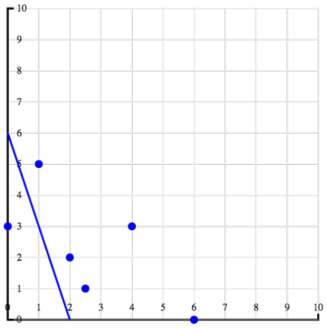


Grading Predictor Functions

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

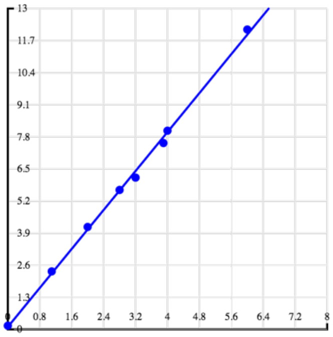


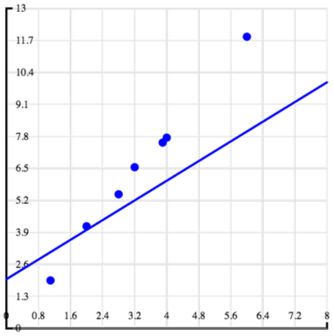


Grade for best predictor:

____

2

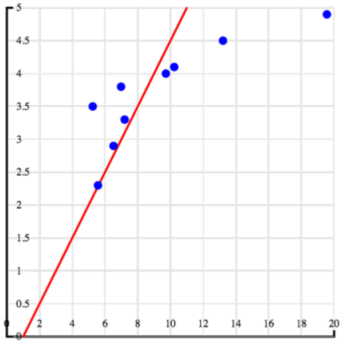


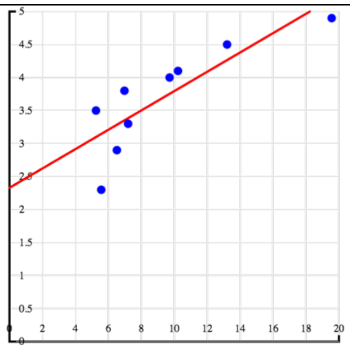


Grade for best predictor:

____

3

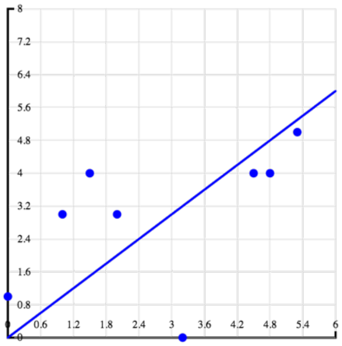


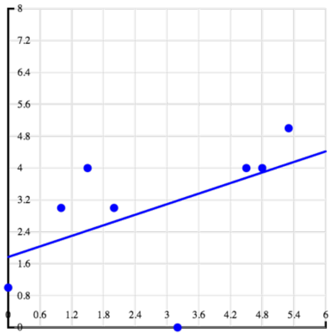


Grade for best predictor:

____

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

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!):

```
select lifespan, name from animals end
```


2. What code will produce the table shown here?

eyes
2
2
8
2

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

```
table1 = select name, legs from animals end
table2 = select legs from mystery end
```

`table2`

Table Plan: Anything Unnecessary?

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 frequency bar chart showing the distribution of `legs`

Are any of the columns unnecessary?

_______myTable-selected_________ =

select ________________________________ **from** _____animals______

end

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

Are any of the columns unnecessary?

___________________________ =

select ________________________________ **from** _____animals______

end

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

Are any of the columns unnecessary?

___________________________ =

select ________________________________ **from** ________________

end

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 the rows need to be in some order?

```
______myTable-ordered________ =

   select _______________________________ from ______animals______
end
```

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

Do the rows need to be in some order?

```
_____________________________ =

   select _______________________________ from ______animals______
end
```

3. We want an alphabetized list of animal names.

Do the rows need to be in some order?

```
_____________________________ =

   select _______________________________ from ________________
end
```

Table Plan: Gross 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	188
Kung Fu Panda 2	665.7	293.5
Man of Steel	668	291
The Sixth Sense	672.8	165.2
Interstellar	675.1	195.3

Do the rows need to be in some order?

___movies-ordered___ = **order** ___movies___:

____

end

Are any of the columns unnecessary?

___total-and-domestic___ = **`select`**

______________________________ **`from`** ______________

`end`

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`

		Total			
Movie Title	**Studio**	**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 the rows need to be in some order?

movies-ordered = **order** movies:

____

end

Are any of the columns unnecessary?

title-and-year = **`select`**

______________________________ **`from`** ______________

`end`

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
`legs <= 4`	
`eyes == 2`	
`legs <> 4`	
`eyes <> 5 - 3`	
`legs == eyes`	

When you finish the first table try these challenge questions:

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

Table Plan: Recent Title and Year

Show the title and year for movies released after 2011, 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 need to get rid of any rows?

____movies-sieved____ = **sieve** ____________ **using** ____________________:

____

end

Do the rows need to be in some order?

____movies-ordered____ = **order** ____________________:

____

end

Are any of the columns unnecessary?

```
______solution4__________ = select

    _______________________________________ from ______________
end
```

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 need to get rid of any rows?

____movies-sieved____ = **sieve** ________________ **using** ____________________________:

____

end

Do the rows need to be in some order?

____movies-ordered____ = **order** ____________________________:

____

end

Are any of the columns unnecessary?

____solution5____ = **select**

______________________________________ **from** ________________

end

Bad Starter Tables!

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

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 need to get rid of any rows?

________________________ = **sieve** _______________ **using** __________________________:

    ____

end

Do the rows need to be in some order?

________________________ = **order** _________________________:

    ____

end

Are any of the columns unnecessary?

________________________ = **select**

    ____ **from** _____________________

end

Unit 9

"Is individual GDP a good predictor of 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 add a column?

______________-extended = **extend** ________________ **using** ________________________:

________________________ : ____

end

Do I need to get rid of any rows?

________________________ = **sieve** ________________ **using** ________________________:

____

end

Do the rows need to be in some order?

________________________ = **order** ________________________:

____

end

Are any of the columns unnecessary?

________________________ = **select**

____ **from** ____________________

end

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 add a column?

____________-extended__ = **extend** ________________ **using** ________________________:

______________________ : ____

end

Do I need to get rid of any rows?

________________________ = **sieve** _______________ **using** __________________________:

____

end

Do the rows need to be in some order?

________________________ = **order** ________________________:

____

end

Are any of the columns unnecessary?

________________________ = **select**

____ **from** ____________________

end

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 add a column?

_____________-extended_ = **extend** ________________ **using** ________________________:

______________________ : ____

end

Do I need to get rid of any rows?

_______________________ = **sieve** _______________ **using** _________________________:

____

end

Do the rows need to be in some order?

_______________________ = **order** ________________________:

____

end

Are any of the columns unnecessary?

_______________________ = **select**

____ **from** ____________________

end

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 add a column?

______________-extended = **extend** ______________ **using** ______________________:

______________________ : ____

end

Do I need to get rid of any rows?

______________________ = **sieve** ______________ **using** ______________________:

____

end

Do the rows need to be in some order?

______________________ = **order** ______________________:

____

end

Are any of the columns unnecessary?

______________________ = **select**

____ **from** ____________________

end

Table Plan

Do I need to add a column?

_____________-extended__ = **extend** _________________ **using** _________________________:

________________________ : ____

end

Do I need to get rid of any rows?

_________________________ = **sieve** ________________ **using** ___________________________:

____

end

Do the rows need to be in some order?

_________________________ = **order** __________________________:

____

end

Are any of the columns unnecessary?

_________________________ = **select**

____ **from** ______________________

end

Table Plan

Do I need to add a column?

____________-extended____ = **extend** __________________ **using** __________________________:

        ________________________ : ____

end

Do I need to get rid of any rows?

_________________________ = **sieve** ________________ **using** ___________________________:

        ____

end

Do the rows need to be in some order?

_________________________ = **order** __________________________:

        ____

end

Are any of the columns unnecessary?

_________________________ = **select**

        ____ **from** ______________________

end

Table Plan

Do I need to add a column?

____________-extended__ = **extend** ______________ **using** ____________________:

______________________ : ____

end

Do I need to get rid of any rows?

______________________ = **sieve** ______________ **using** ____________________:

____

end

Do the rows need to be in some order?

______________________ = **order** ______________________:

____

end

Are any of the columns unnecessary?

______________________ = **select**

____ **from** ____________________

end

Query Reference

Select

What it's for:

```
select column1 , column2 , column3 from table end
```

Order

What it's for:

```
order table :
    column1 ascending ,
    column2 descending
end
```

Sieve

What it's for:

```
sieve table using column2 :
    column2  > 42
end
```

Extend

What it's for:

```
extend table using column1 , column2 :
    new-column1 : (2 * column1) – column2
    new-column2 : column2 / 4
end
```

Contracts

Name	Domain	Range	Example
num-max			num-max(-1, 3)
string-length			string-length("pyret")
string-repeat	String Number	String	