Data 8 Spring 2017 - Midterm 1 Review Worksheet II

I. Tables

You have two tables, animals and types. Use the two tables to answer the questions.

animals

Class	Famous?	Name	Specific Type
Asteroidea	True	Patrick Star	Starfish
Mammalia	True	Simba	Lion
Mammalia	False	McGruff	Bloodhound
Reptilia	True	Geico Gecko	Day Gecko

(87 more rows)

types

Specific Type	Weight (lbs)	Color
Starfish	0.125	Pink
Lion	500	Gold
Orca	2000	Multicolored

(102 more rows)

Write the code that you would use to show the following:

1. What is the name of the heaviest animal in the animal table?

2. Create a new column for types with the weight in kilograms (1 pound = 2.2 kilograms).

3. Figure out the average weight for each class in the animals table.

4. Create an array of the types of animals that are multicolored.

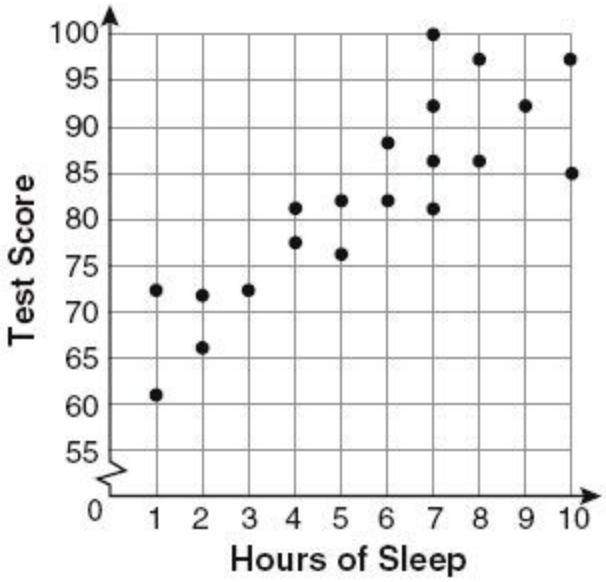
5. True or False: Patrick Star is heavier than the Geico Gecko.

6. How many animals in the animals table are in the reptilia class?

7. Create a table that has the counts for how many famous and not famous animals there are in each class for animals.

II. Plot

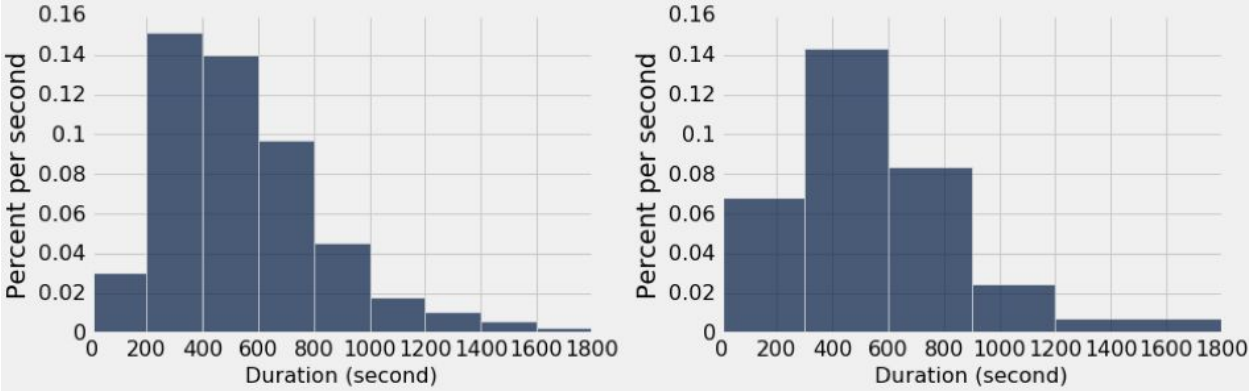
Look at the following scatter plot. What can you say about the relationship between the amount of sleep that a student gets the night before a test and their score?



III. Histograms

(From Spring 2016 Midterm)

The two histograms of bike trip durations below were both generated by trip.hist(...) using different bins (same data, different bins).



Write the proportion of trips that fall into each range of durations below. Show your work. If it is not possible to tell from the histograms, instead write NOT ENOUGH INFO.

1. Between 200 (inclusive) and 400 (exclusive) seconds

2. Between 300 (inclusive) and 900 (exclusive) seconds

3. Between 400 (inclusive) and 900 (exclusive) seconds

4. Between 200 (inclusive) and 300 (exclusive) seconds

IV. Functions

1. Define a function that when given an array of numbers, prints all of the even numbers in the array.

2. Define a function that when given a table and a column name, returns an array of all the different possible values in that column.

3. Using a for loop, simulate sampling randomly 1000 times a number from 0 to 10 and output the percentage of number 7.

4. Fill in the blanks so that the following code outputs 2017.

```
big_data = make_array(987654321, 924645644, 54500654, 8506006)
# Note that 987654321 - 924645644 - 54500654 - 8506006 = 2017

difference = _____
for index in np.arange(_____):
    difference = abs(big_data.item(_____) - _____)
print(difference)
```

V. Probabilities

(From Fall 2016 Probability Review)

We draw 2 tickets at random, **with replacement**, from a box with the tickets described below.

Number of Tickets	**Red**	**Blue**	**Green**
Smooth	20	15	15
Jagged	30	15	5

1. What is the probability that both tickets are green?

2. What is the probability that both tickets are the same color?

3. What is the probability that both tickets are blue and jagged?

(From Spring 2016 Midterm)
A study followed 369 people with cardiovascular disease, randomly selected from hospital patients. A year later, those who owned a dog were four times more likely to be alive than those who didn't.

1. Circle True or False: This study is a randomized controlled experiment.

2. Circle True or False: This study shows that dog owners live longer than cat owners on average.

3. Circle True or False: This study shows that for someone with cardiovascular disease, adopting a dog will probably cause them to live longer.

Table Functions and Methods

Table	Create an empty table, usually to extend with data
with_columns	Create a copy of a table with more columns
column	Create an array containing the elements of a column
num_rows	Compute the number of rows in a table
num_columns	Compute the number of columns in a table
labels	Lists the column labels in a table
select	Create a copy of a table with only some of the columns
drop	Create a copy of a table without some of the columns
relabel	Modifies the existing table in place, changing the column heading in the first argument to the second
relabeled	Returns a new table with the column heading in the first argument changed to the second
sort	Create a copy of a table sorted by the values in a column. Defaults to ascending order unless "descending = True" is included
where	Create a copy of a table with only the rows that match some predicate

take	Create a copy of the table with only the rows whose indices are in the given array
barh	Draws a bar chart of the frequencies of a categorical distribution
histogram	Draws a histogram of a numerical distribution
apply	Returns an array of a a function applied to some row in a table
group	Create a copy of the table with all rows with the same values in a certain column aggregated into one row in the new table
pivot	Create a copy of the table with a column for each element in the first argument and a row for each element in the second argument and aggregates values
join	Create a copy of the table that is the result of joining the columns of two tables, with a row for each shared value in the two tables
sample	Draws some number of rows at random from a table. By default, with replacement.
proportion_from_distribution	Returns a new table with an additional column corresponding to proportions from a random sample of some size of the proportions in a column.

Array Functions and Methods

max	Returns the maximum value of an array
min	Returns the minimum value of an array
sum	Returns the sum of the values in an array
len	Returns the length (number of elements) of an array
make_array	Makes a numpy array with the values passed in
np.average	Returns the mean value of an array
np.diff	Returns a new array of size len(arr)-1 with elements equal to the difference between adjacent elements
np.sqrt	Returns an array with the square root of each element
np.arange	Returns an array of an end-exclusive range of variable step size
arr.item	Returns the i-th item in an array (remember Python indices start at 0!)
np.random.choice	Picks one (by default) or some number 'n' of items from an array at random. By default, with replacement.
np.count_nonzero	Returns the number of non-zero (or True) elements in an array.
np.append	Returns a copy of the input array with some item (must be the same type as the other entries in the array) appended to the end.