Data 8 Spring 2021
Discussion: Functions and Table Methods (Project 1 Discussion)

Welcome to Project 1 Discussion! This week we will be discussing functions and table methods like group. The group function allows us to aggregate the unique entries in one or more columns.

Question 1. Fun with Functions

    a. The following code has a number of errors in it. Which ones can you identify? ①

```
def hypotenuse(a, b):
        ```Returns the length of the hypotenuse of a right triangle,
        the squareroot of a squared + b squared```
        squares = make_array(a, b)*2   ② ③ ④ ⑤
        sum = sum(squares)          summed
        squareroot = np.sqrt(sum)   → summed
        print(squareroot)  return  Square Root
    A = 5
    B = 5
    C = squareroot  hypotenuse (A, B)   ⑥
```

(handwritten annotations: "summed", "squareroot" circled, "summed", "return Square Root", "hypotenuse (A, B)")

    b. Write a function that takes in one argument, a table `tbl`, another argument, a name of a column in that table `col`, and a boolean `largest`, and returns a table that contains the rows that have the ten largest or ten smallest values for the specified column, largest if the boolean `largest` is `True`, smallest if the boolean argument is `False`.

```
def top_ten(tbl, col, largest):
    sorted_tbl = tbl.sort ( col, descending = largest)
    ten_rows = sorted_tbl. take (np.arange (10))
    return  ten_rows
```

    c. Can a function take no arguments? When would you use a function with no arguments? How do you call a function without arguments? How does that compare to using a function as an argument?

```
def no_args():
    return 'No args'
```
→ c = no_args ()
    ↳ 'no args'

```
def args (x):
    return x
```
→ b = args (10)
    ↳ 10

Question 2. Ian has opened up a chocolate store where he sells small boxes of chocolates in groups of different sizes and colors. His table `chocolates` is as follows:

Color	Shape	Amount	Price ($)
Dark	Round	4	1.30
Milk	Rectangular	6	1.20
White	Rectangular	12	2.00
Dark	Round	7	1.75
Milk	Rectangular	9	1.40
Milk	Round	2	1.00

Notice that the table contains multiple rows containing information about chocolates of the same color. We would like to figure out how many chocolates of each color he has for sale in total, and what the cost would be to purchase all chocolates of each unique color.

a. Write a line of code that will return a new table which displays the total number of boxes for each color.

chocolates . group ('color') →
Dark   2
Milk   3
White   1

b. Write a line of code which will return a new table with the total number of chocolates and the total cost for each unique color. For example, the row for "Dark" should have a total of 4+7=11 chocolates, and a total cost of $1.30 + $1.75 = 3.05.

chocolates. group ('color', sum). drop ('Shape sum')

Question 3. Some rows from the table `ca` are shown below. The table contains information about the most common baby names in California and the number of those occurrences in a particular year, from the years 1910-2019. (This dataset was submitted by a fellow Data 8 student!)

State	Sex	Year	Name	Occurrence
CA	F	1910	Mary	295
CA	F	1910	Helen	239
CA	F	1910	Dorothy	220
CA	F	1910	Margaret	163
CA	F	1911	Mary	10

a. Write a line of code that will return the most popular name over all the years.
   Hint: Think about how to use the second argument in `.group`

ca. group ('Name', sum). sort ('occurrence sum', descending = True). column ('Name'). item (0)

b. Instead of the most popular name over all the years, write a line of code that will return the top 10 most popular names over all the years.

where ('Sex', are. equal- to ('F'))

↕ same

c. The top 10 names all appeared to be male names. Write a line of code that would return the most popular female names instead.

ca. where ('Sex', 'F'). group ('Name', sum). Sort ('occurrence sum', descending = True).take (np. orange (10)). Column ('Name')

d. Write a line of code that will return the most popular female name in 1969

e. Write a function `most_popular_female_name` that takes in a year as an argument and returns the most popular female name in that year.

```
def most_popular_female_name(year):
```
return   ca. where ('Sex', 'F'). where ('Year', year). Sort ('occurrence', descending = True). Column ('Name') item (0)

f. The `ca` table is from 1920-2019. Define the years table with a column year and a row for each year from 1910-2019 (inclusive). Then create the table `popular_female_names` that has 2 columns, a year and a column for the female name that is most popular.

```
years = 
```
Table(). with - column ('Year', np. arange (1910, 2020))
```
most_popular_female_names_array = 
```
years. apply ('Year', most - pop_fem. name)
```
popular_female_names = 
```
year. with - column (' pop fem , most - popular - name'   female -names - array)

g. Write a line of code that will generate the following bar chart:
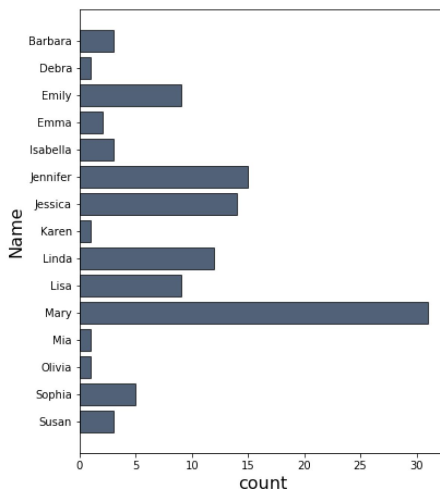
CAN'T DO THIS WEEK

Question 4. The table below, called `weights`, contains information about the weights of the chocolates that are sold. The weights of the chocolates differ depending on the shape, and round chocolates have two different sizes.

Shape	Weight(g)
Round	3.1
Round	4.25
Rectangular	3.6
Triangular	2.9

The following line of code has been executed in a blank cell. Take a moment to discuss with your neighbors what the resulting table will look like. Then, write the number of columns and rows in the resulting table, and describe the information in the table in 1-2 sentences.

Hint: It may help to draw a sketch of the resulting table!

```
chocolates.join('Shape', weights)
```

Question 5. We will continue with the same table as before, copied below for your convenience.

Color	Shape	Amount	Price ($)
Dark	Round	4	1.30
Milk	Rectangular	6	1.20
White	Rectangular	12	2.00
Dark	Round	7	1.75
Milk	Rectangular	9	1.40
Milk	Round	2	1.00

Write code to create a pivot table on the colors and shapes of chocolates, finding the average price for each color-shape combination. Then, fill in the blank table in the image of the resulting table.

Hint: You can use the `np.average` function to find the average of an array of inputs. The average of no values is marked as zero.

	Rectangular	Round
White		
Milk		
Dark		