HW 04 Written Work

● Graded

Student

Sangwon Ji

Total Points

40 / 40 pts

Question 1

Q1.2

8 / 8 pts

✔ **+ 8 pts** Correct plot obtained by:

yelp_and_google = burritos.select(“Yelp”, “Google”)
yelp_and_google.scatter(“Yelp”, “Google”)
OR
burritos.scatter("Yelp", "Google")

+ 0 pts Incorrect/blank

Question 2

Q1.3

8 / 8 pts

✔ **+ 8 pts** Reasonable explanation (Google ratings are consistently higher, positive association, etc)

+ 0 pts Click here to replace this description.

Question 3

Q1.6

8 / 8 pts

✔ **+ 8 pts** Correct plot obtained by:

burritos.hist(“Cost”, bins = bins)

NOTE: the student has to include the bins in order to get credit

+ 0 pts Click here to replace this description.

Question 4

Q2.2

8 / 8 pts

✔ **+ 4 pts** Correct Answer: The arrays would be different.

✔ **+ 4 pts** Correct explanation: .group preserves row order, so item order in array would change.

+ 0 pts Blank / Incorrect

Question 5

Q2.4

8 / 8 pts

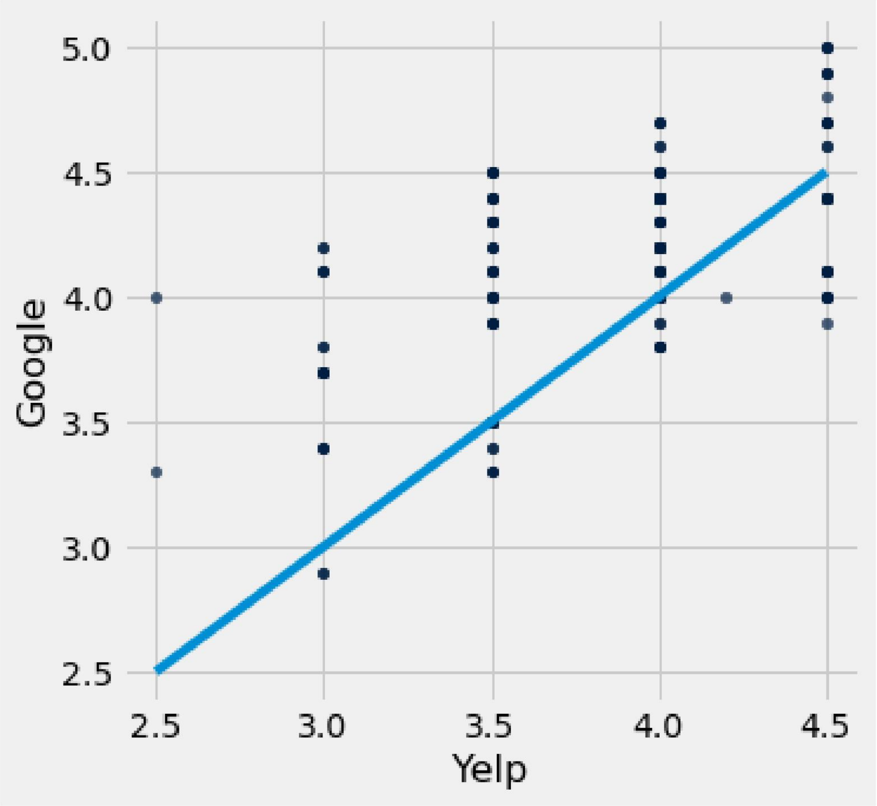
✔ **+ 8 pts** Any reasonable explanation: the range is 0 because there is only one person, there was missing data, the range is actually 0, etc

+ 0 pts Click here to replace this description.

Question 2. Let's look at how the Yelp scores compare to the Google scores in the `burritos` table. First, assign `yelp_and_google` to a table only containing the columns `Yelp` and `Google`. Then, make a scatter plot with Yelp scores on the x-axis and the Google scores on the y-axis. **(8 Points)**

```
In [40]: yelp_and_google = burritos.select('Yelp','Google')
         yelp_and_google.scatter('Yelp','Google')

         # Don't change/edit/remove the following line.
         # To help you make conclusions, we have plotted a straight line on the graph (y=x).
         plt.plot(np.arange(2.5,5,.5), np.arange(2.5,5,.5));
```



```
In [41]: grader.check("q1_2")
```

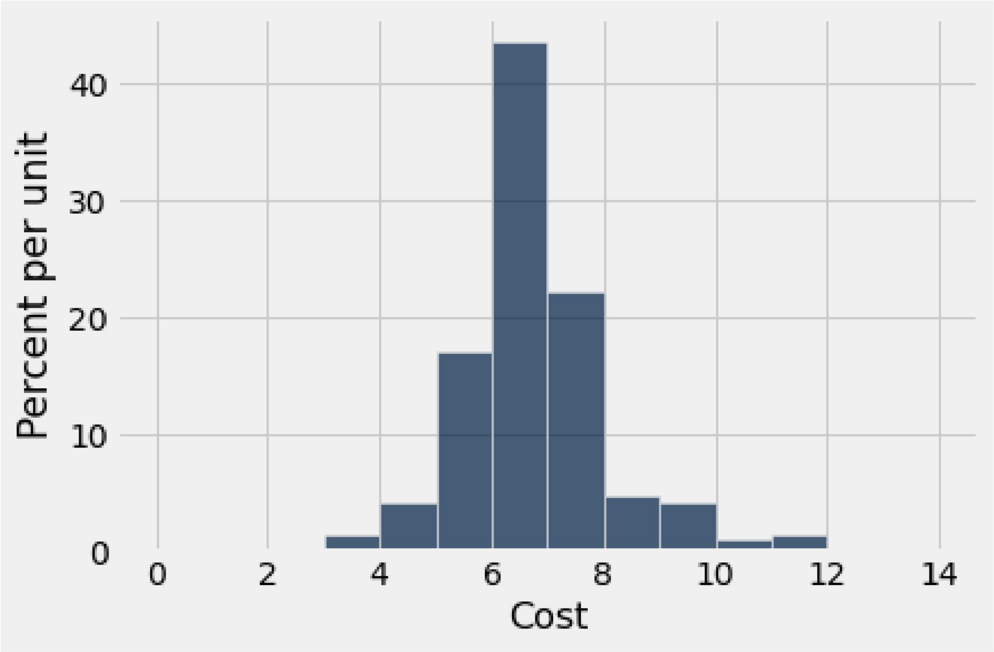
```
Out[41]: q1_2 results: All test cases passed!
```

Question 3. Looking at the scatter plot you just made in Question 1.2, do you notice any pattern(s) (i.e. is one of the two types of scores consistently higher than the other one)? If so, describe them **briefly** in the cell below. **(8 Points)**

Not all of them, but mostly, the scores seems to come out higher in the Google's score more. With the scatter plot above, it is easy to find out that looking at the x-axis, Yelp scores and comparing it with the scores that come out on Google, Google's ones are higher. Also, with the blue line that shows the equal number for Yelp and Google, majority of the scores seems to be above. This seems to be a consisten pattern as numbers doesn't go that below to the ones of Yelp ones.

Question 6. Mira thinks that burritos in San Diego are cheaper (and taste better) than the burritos in Berkeley. Plot a histogram that visualizes that distribution of the costs of the burritos from San Diego in the `burritos` table. Also use the provided `bins` variable when making your histogram, so that the histogram is more visually informative. **(8 Points)**

```
In [47]: bins = np.arange(0, 15, 1)
         # Please also use the provided bins
         burritos.hist("Cost", bins = bins)
```



Question 2. At the moment, the `Job` column of the `sf` table is not sorted (no particular order). Would the arrays you generated in the `Jobs` column of the previous question be the same if we had sorted alphabetically instead before generating them? Explain your answer. To receive full credit, your answer should reference *how* the `.group` method works, and how sorting the `Jobs` column would affect this. **(8 Points)**

Note: Two arrays are the **same** if they contain the same number of elements and the elements located at corresponding indexes in the two arrays are identical. An example of arrays that are NOT the same: `array([1,2]) != array([2,1])`.

The results won't come out the same as it will display different result since the whole array will also come out differently. .group works by collecting and rowing the value in a column by using values and combinations. Sorting the Jobs column will affect this as it is located differntly and would change the rows, eventually showing the different results.

Question 4. Give an explanation as to why some of the row values are `0` in the `department_ranges` table from the previous question. **(8 Points)**

It is becuase there isn't number that's existing for that department in the organization group. There is no data for it as one department belongs to one organization group. There is why the pivot is outputing the data 0.