HW 03 Written Work

● Graded

Student

Sangwon Ji

Total Points

26 / 30 pts

Question 1

Q2.5 Resolved **0** / 4 pts

+ 4 pts Correct scatterplot; birth rate vs. death rate and base column is 2015

✔ **+ 0 pts** Incorrect/blank

The question wanted the association between birth and death rates, not birth and death count.

Regrade Request **Submitted on: Jul 14**

I think it shows the association between birth and death rates in the scatter plot that I've created

Looking at your code, your birth_rates and death_rates variables will evaluate to the correct array. However, you didn't plot that in your scatter plot. What you did plot is the column "BIRTHS" and "DEATHS", columns from the pop table that do not represent their respective rates.

Reviewed on: Jul 15

Question 2

Q3.1 **4** / 4 pts

✔ **+ 4 pts** Correct plot of Boston table

+ 0 pts Incorrect/blank

Question 3

Q3.2 **4** / 4 pts

✔ **+ 4 pts** Correct plot of Manila table

+ 0 pts Incorrect/blank

Question 4

Q3.6 **4** / 4 pts

✔ **+ 2 pts** Addressed first part sufficiently: *Identify one difference between the histograms*

✔ **+ 2 pts** Addressed second part sufficiently: *Comment on the average and/or skew of each histogram*

+ 0 pts Incorrect/blank

Question 5

Q3.7 **4** / 4 pts

✔ **+ 4 pts** Any valid explanation on why ride times in Manila are longer than in Boston

+ 0 pts Click here to replace this description.

Question 6

Q4.2 **5** / 5 pts

✔ **+ 2.5 pts** Histogram C

✔ **+ 2.5 pts** Correct explanation; can discuss range, gaps, distribution of points, etc

+ 0 pts Incorrect/blank

Question 7

Q4.4 **5** / 5 pts

✔ **+ 2.5 pts** Histogram B

✔ **+ 2.5 pts** Correct explanation; can discuss range, gaps, distribution of points, etc

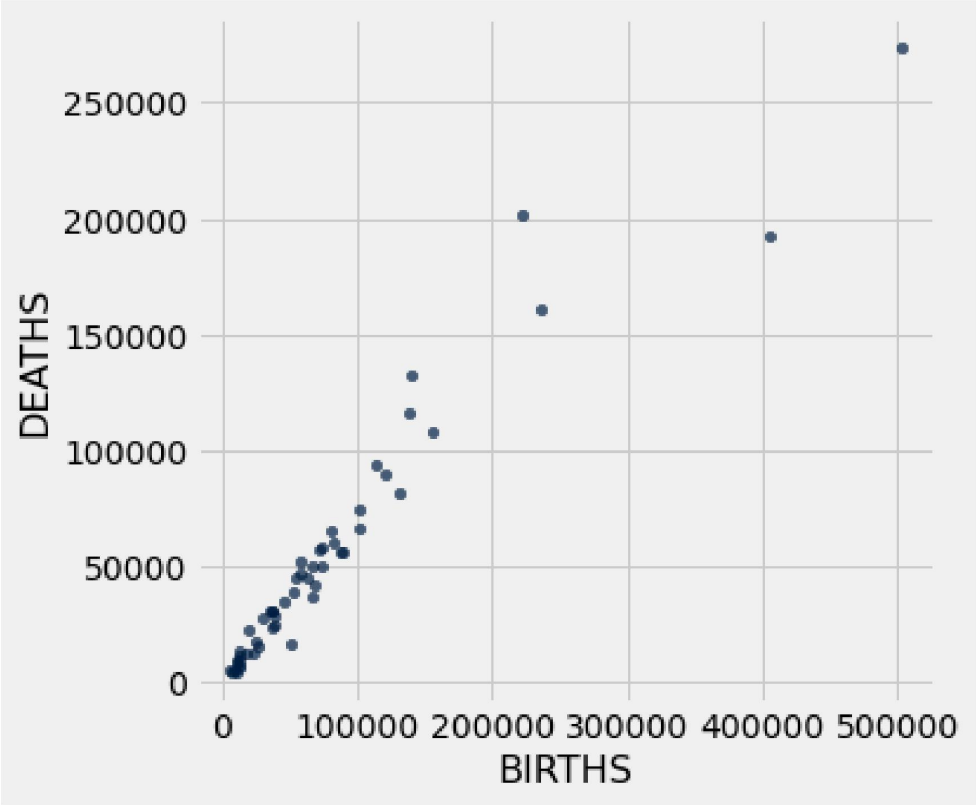
+ 0 pts Incorrect/Blank

Question 5. In the code cell below, create a visualization that will help us determine if there is an association between birth rate and death rate during this time interval. It may be helpful to create an intermediate table here. **(4 Points)**

Things to consider:

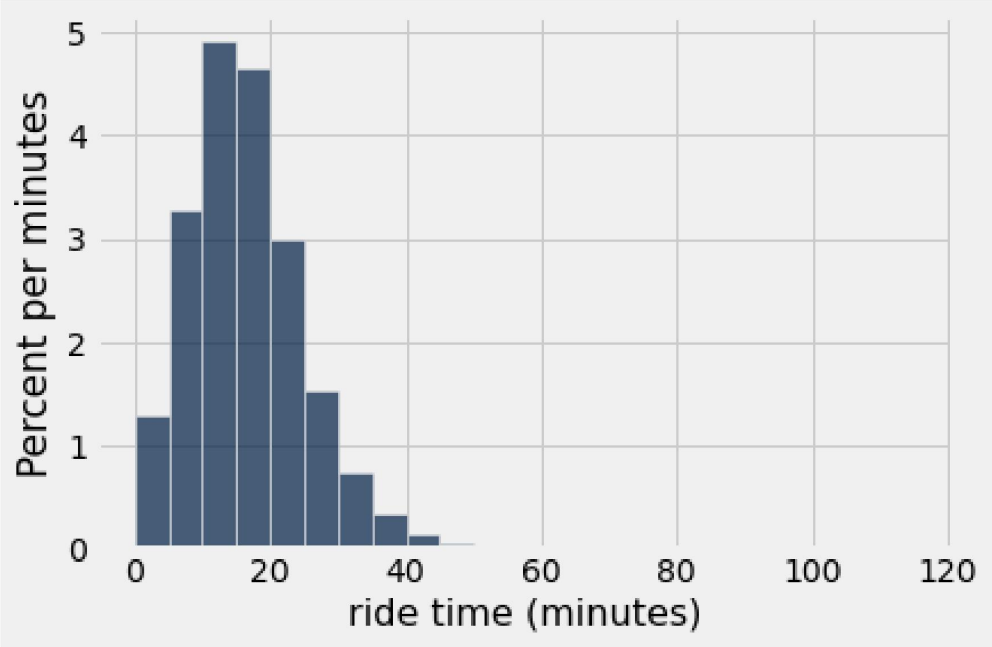
- What type of chart will help us illustrate an association between 2 variables?
- How can you manipulate a certain table to help generate your chart?
- Check out the Recommended Reading for this homework!

In [28]:
```
# In this cell, use birth_rates and death_rates to generate your visualization
birth_rates = pop.column('BIRTHS') / pop.column('2015')
death_rates = pop.column('DEATHS') / pop.column('2015')
pop.scatter("BIRTHS", "DEATHS")
```



Question 1. Produce a histogram that visualizes the distributions of all ride times in Boston using the given bins in `equal_bins`. **(4 Points)**

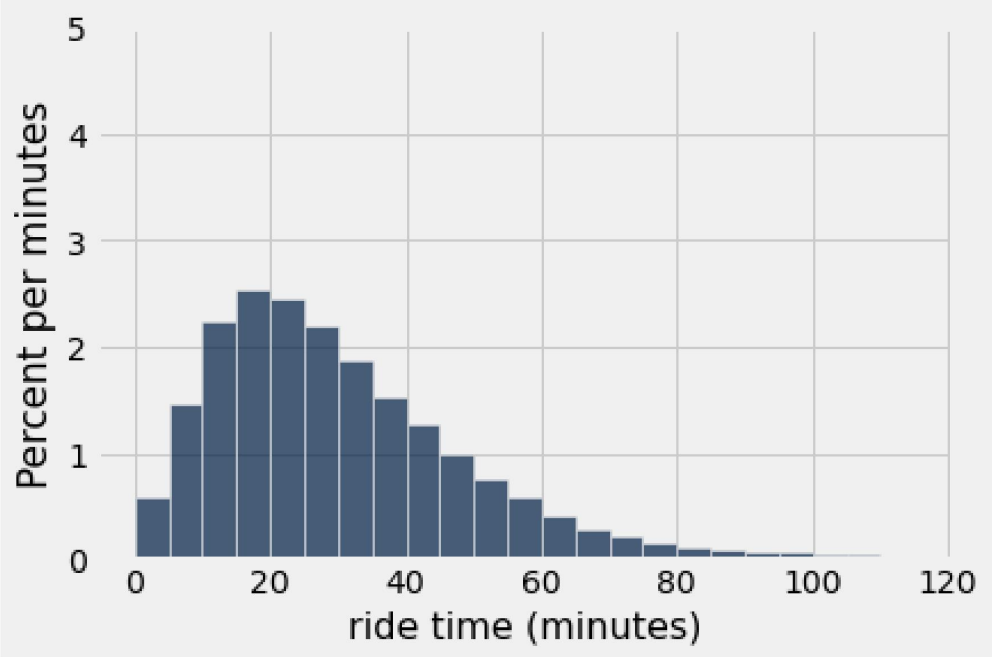
```
In [33]: equal_bins = np.arange(0, 120, 5)
         boston.hist('ride time', bins=np.arange(0, 120, 5), unit='minutes')
```



Question 2. Now, produce a histogram that visualizes the distribution of all ride times in Manila using the given bins. **(4 Points)**

```
In [34]: equal_bins = np.arange(0, 120, 5)
         manila.hist('ride time', bins=np.arange(0, 120, 5), unit='minutes')

         # Don't delete the following line!
         plots.ylim(0, 0.05);
```



Question 6. Identify one difference between the histograms, in terms of the statistical properties. Can you comment on the average and/or skew of each histogram? **(4 Points)**

Hint: The best way to do this is to compare the two histograms (from 3.1 and 3.2) visually.

Looking at two histograms visually, first distinctive feature is that the variables of histogram of manila is right-skewed, continuing numbers but more concentrated on the first 30 minutes, and the histogram of boston has no variables of ride time that goes over 50 minutes, just concentrated on the first minutes, with the number spiked at 10 to 20 minutes. The exact number of average is really hard to define just looking at the histogram, but we are able to get a approximate estimation of it. The width of the bins are defined, and the numbers, percent per minutes are not exact but you can tell how much is it. However, as the skewness exists, the average of the data could come out different. And the basic difference is where it's more concentrated in and has different values of numbers, and how one's continuous while the other one doesn't.

Question 7. Why is your solution in Question 6 the case? Based on one of the following two readings, why are the distributions for Boston and Manila different? **(4 Points)**

- Boston reading
- Manila reading

Hint: Try thinking about external factors of the two cities that may be causing the difference! There may be multiple different factors that come into play.

Manila seems to have a bad traffic, so regardless of the miles that actually run, even with the similar distance of the arrival, it might actually take more time than those of boston. According to the reading, Manila seems to have the worst traffic in the world. It is indicated that Manila has inefficient design of the road and the law that is fragile, it causes a chaos in the road making the drive more longer than usual. That might be one of the reason that's causing the difference to having experience in Boston. Added to that, there might be lot of external reason that come into play to make the difference.

Question 2. State at least one reason why you chose the histogram from Question 1. **Make sure to clearly indicate which histogram you selected** (ex: "I chose histogram A because ..."). **(5 Points)**

I chose 3. Histogram C because I could clearly see that the range of x varies from -2 to 3, with similar to Histogram A but C has more concentrated on the number between -1 and 0 as the scatterplot also concentrates more highly on those numbers. Additionally, how is continuing to the end well represents the scatterplot as well.

Question 4. State at least one reason why you chose the histogram from Question 3. **Make sure to clearly indicate which histogram you selected** (ex: "I chose histogram A because ..."). **(5 Points)**

I chose 2. Histogram B becuase I could clearly see that the range of y varies from -1.5 to -0.5 and 0.5 to 1.5, and it's the only histogram that corresponds to those of the y-value.