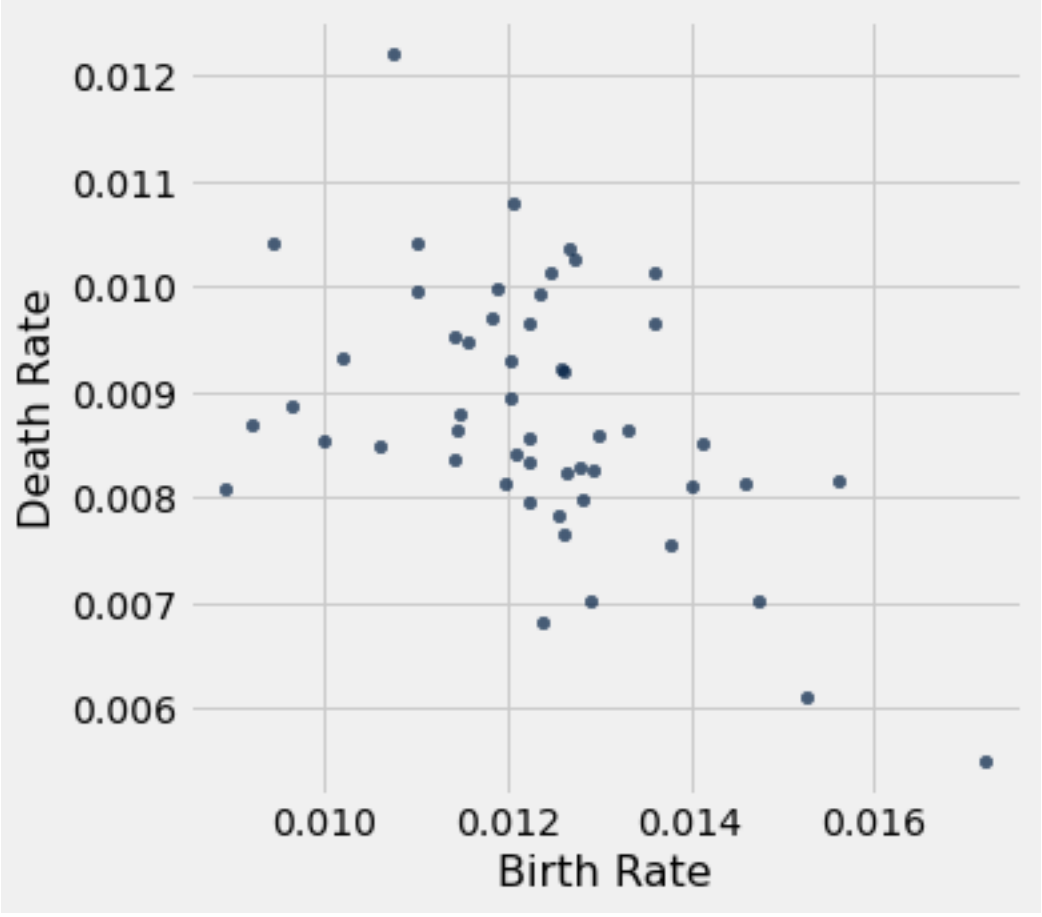
Question 5. In the code cell below, create a visualization that will help us determine if there is an association between birth rate and death rate during this time interval. It may be helpful to create an intermediate table here. **(4 Points)**

Things to consider:

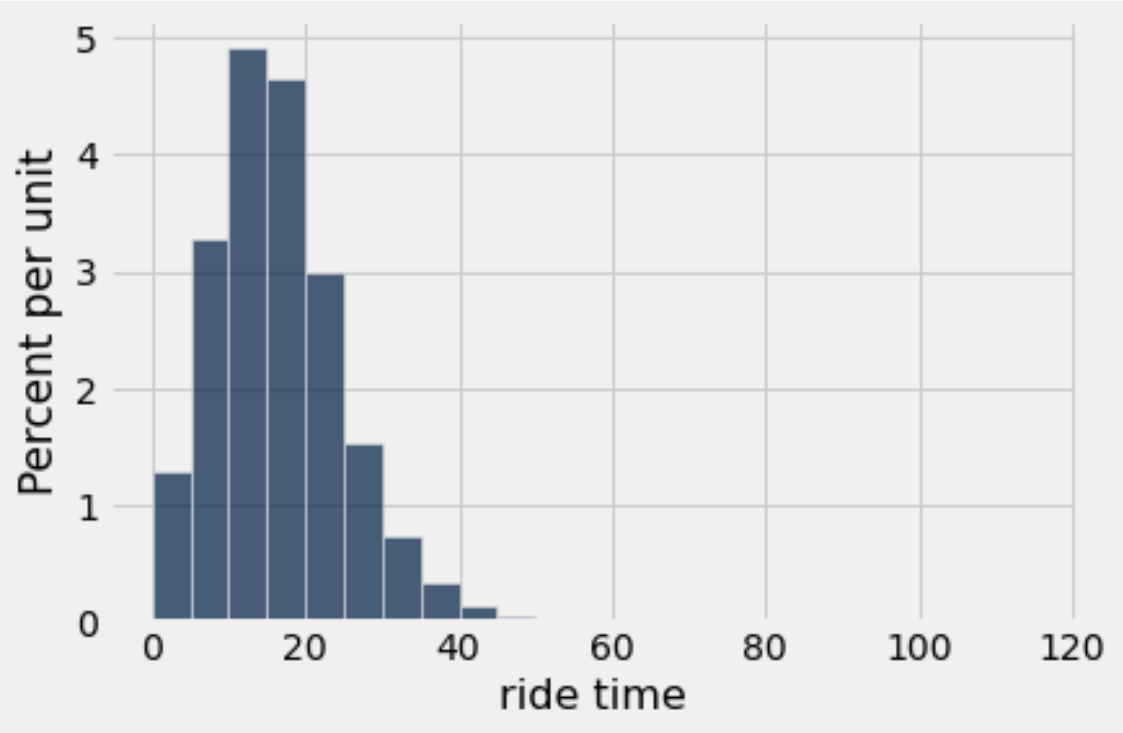
- What type of chart will help us illustrate an association between 2 variables?
- How can you manipulate a certain table to help generate your chart?
- Check out the Recommended Reading for this homework!

In [39]:
```
# 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.with_columns("Birth Rate", birth_rates, "Death Rate", death_rates).scatter(8, 9) # SOLUTIO
```



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

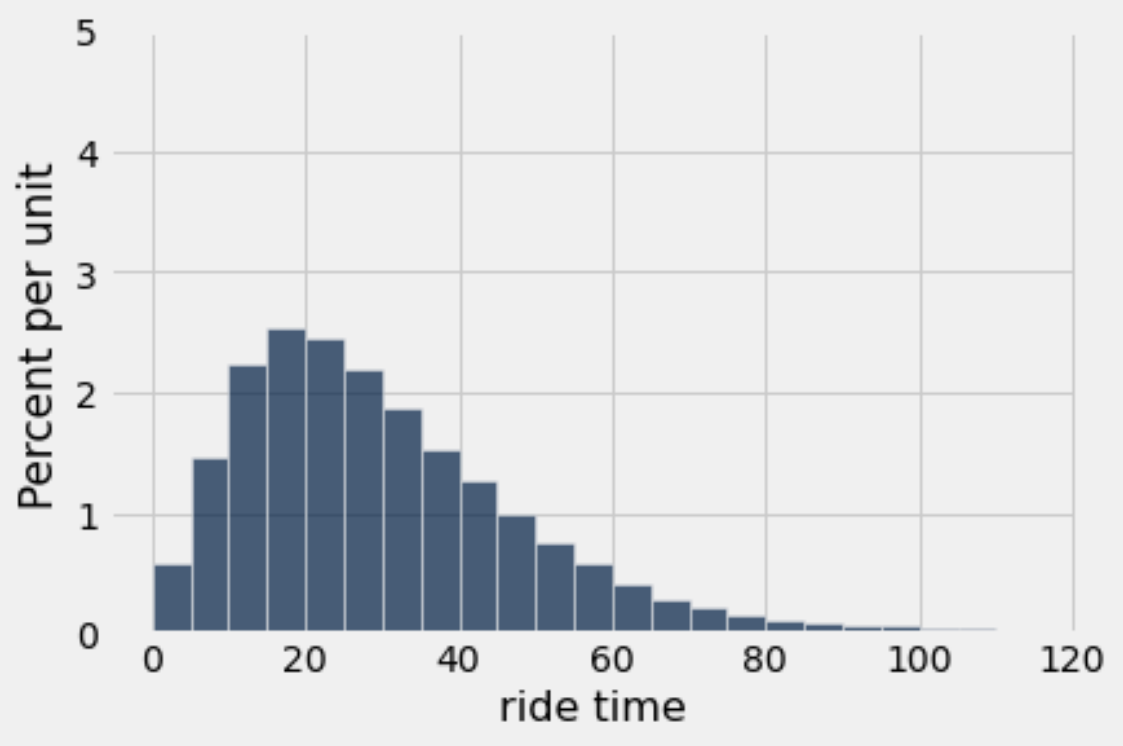
```
In [44]: equal_bins = np.arange(0, 120, 5)
         boston.select("ride time").hist(bins = equal_bins) #SOLUTION
```



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

```
In [45]: equal_bins = np.arange(0, 120, 5)
         manila.select("ride time").hist(bins = equal_bins) #SOLUTION

         # 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.

Type your answer here, replacing this text.

SOLUTION: Long rides make up a greater proportion of all Uber rides in Manila than in Boston. The Manila histogram has a long right tail (right skewed).

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.

Type your answer here, replacing this text.

SOLUTION: This could be because there's more traffic in Manila, or because the weather is not pleasant in Boston in the winter, so people may choose to take a car for short trips rather than walk.

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)**

Type your answer here, replacing this text.

SOLUTION: Because there are no gaps in the X-variable, we would expect the histogram for X to have no gaps in it. Also, because the two masses on the scatter plot overlap in the area between -1 and 0, we would expect there to be more mass in the -1 to 0 area of the histogram, since each vertical slice in this range contains more points. Also, the values of the X-variable range from -2 to 2, which fits the range of values in histogram C.

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)**

Type your answer here, replacing this text.

SOLUTION: There is a gap in the points in the Y-direction, so we would expect a gap in the histogram of those values. Also, the range of values covered by the Y-variable range from -1.5 to 1.5, which fits the range of values in histogram B.