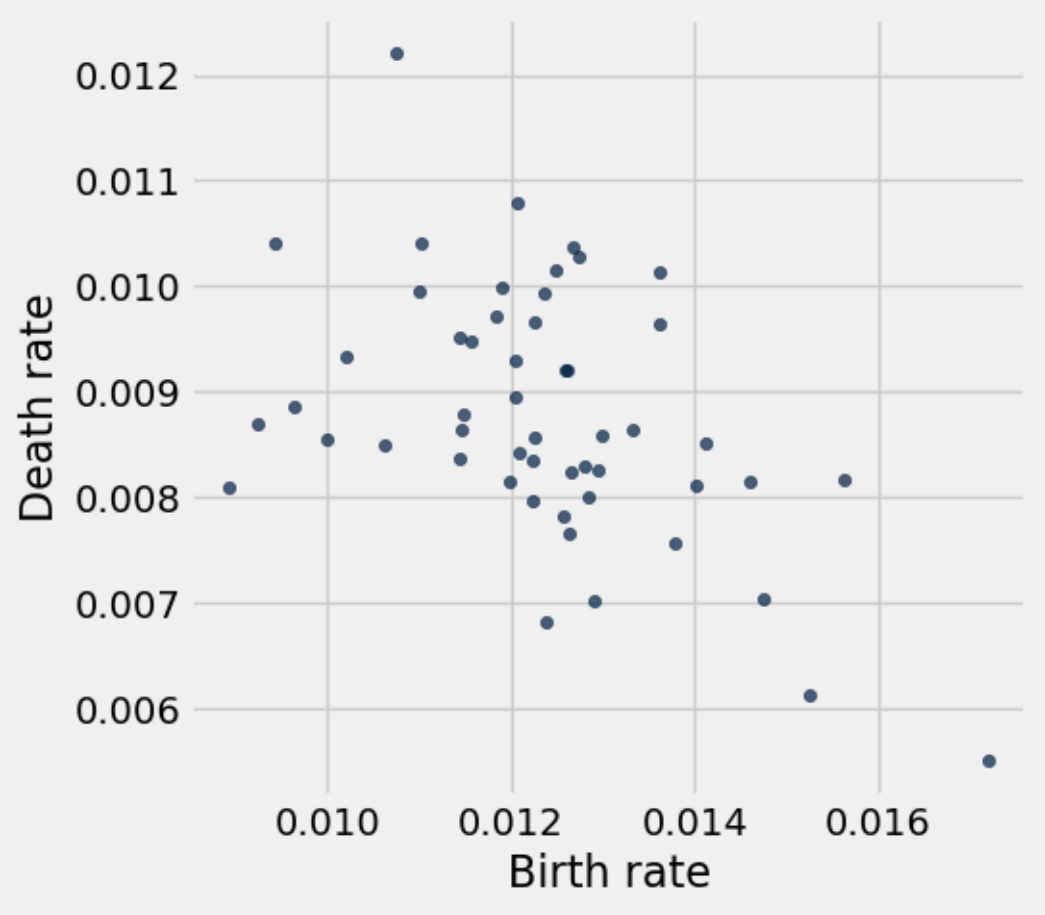
Question 5. In the code cell below, create a visualization that will help us determine if there is an association between birth rate and death rate during this time interval. It may be helpful to create an intermediate table here. **(4 Points)**

Things to consider:

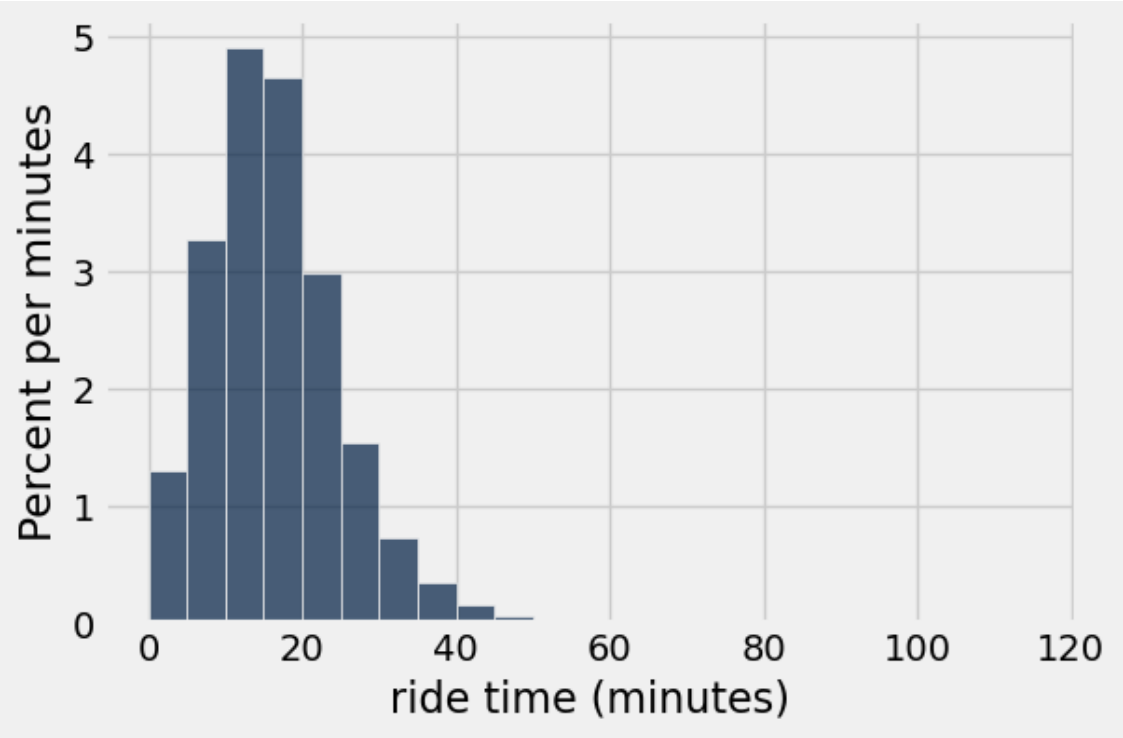
- What type of chart will help us illustrate an association between 2 variables?
- How can you manipulate a certain table to help generate your chart?
- Check out the Recommended Reading for this homework!

```
In [17]: # 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')
         birth_vs_death = Table().with_columns(
             "Birth rate", birth_rates,
             "Death rate", death_rates
         )
         birth_vs_death.scatter("Birth rate", "Death rate")
```



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

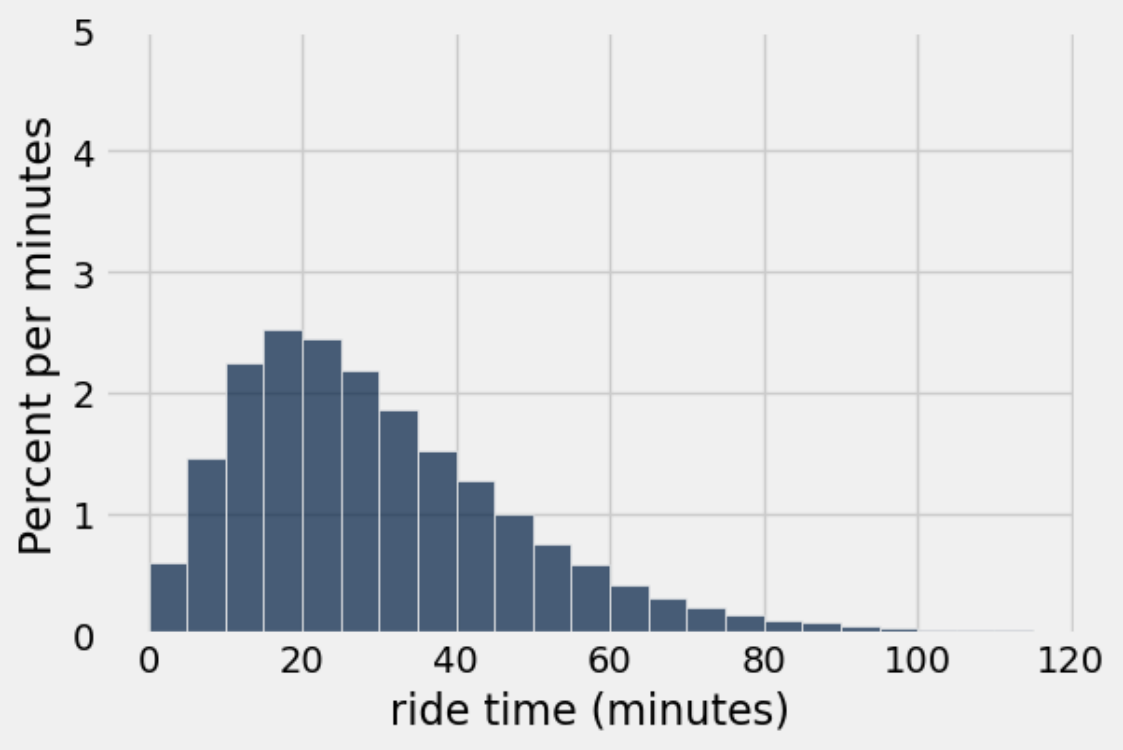
```
In [48]: equal_bins = np.arange(0, 120, 5)
         boston.hist("ride time", unit="minutes", bins=equal_bins)
```



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

```
In [49]: equal_bins = np.arange(0, 120, 5)
         manila.hist("ride time", unit="minutes", bins=equal_bins)

         # 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.

The Manila histogram is visibly more skewed to the right than the Boston histogram, and it has a greater range of ride times. On the other hand, the Boston histogram is more strongly centered around 14-20 minutes and is less varied than the Manila histogram.

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.

According to the article, traffic in Manila is bad primarily due to a lack of enforcement and a disregard for laws. Because of the increased traffic, Uber rides could take much longer than they normally would in places with better traffic. This could be the cause of the wider range in the Manila histogram.

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 histogram C. Looking at just the x-positions of the points on the scatterplot, we see that there is overlap between -1 and 0 between the two clusters, and there are many points farther out to the right. This would cause the histogram to be skewed to the right. Also, we see a few points with an x-coordinate of less than -2. The histogram that best matches this is 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)**

I chose histogram B. Looking at the scatterplot, we see that there aren't any points with a y-coordinate near 0. The only histogram with a gap at 0 is histogram B.