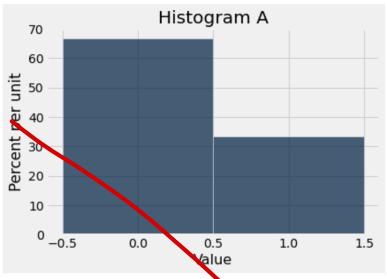

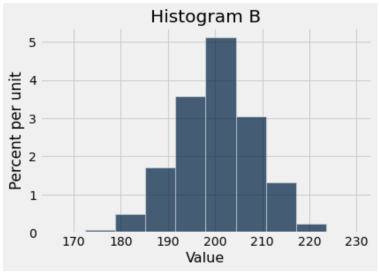

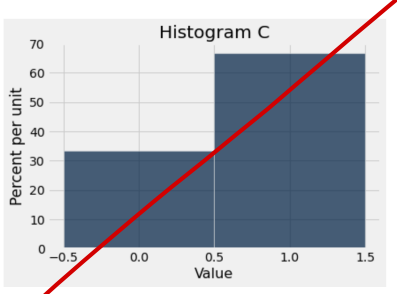

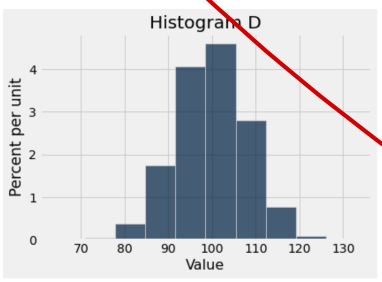

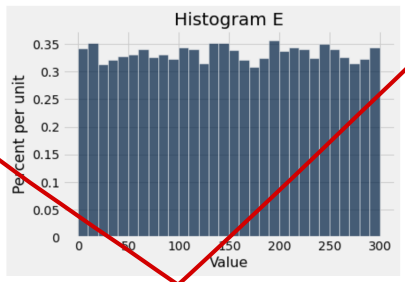

(a) (3 pt) Which histogram is produced by the code? **Fill in exactly one bubble.**

○ A ○ B ○ C ○ D ○ E

(b) (3 pt) In the box below, briefly explain your choice of histogram.

5. (6 points) Song Durations

Christina is interested in learning more about the duration of songs on Spotify. She collects a random sample of 400 songs listed on the platform and stores the data in the table `songs`, which has one column labelled `"Duration"`. The average song duration in the sample is 185 seconds and the standard deviation is 25 seconds. Christina wants to use this sample of songs to make some estimates about the population of songs and their durations. ”

(a) (3 pt) Define a function `song_ci` that constructs a 95% confidence interval for the population mean as follows and returns it as an array. The function takes in the argument `reps`, the number of bootstrap repetitions wanted.

```
def song_ci(reps):

    stats = _____

    for _____:

        resample = _____

        new_mean = _____

        stats = _____

    left_end = _____

    right_end = _____

    return _____
```

(b) (3 pt) Christina creates an interval by using `song_ci(10000)`. To get a more accurate estimate at the same level of confidence, Christina would like to create a new 95% confidence interval that is half as wide as this one. Which one of the following do you think is the best advice for her? **Fill in exactly one bubble.**

- ◯ She should use `song_ci(20000)`
- ◯ She should use a sample of size 800
- ◯ She should use `song_ci(40000)`
- ◯ She should use a sample of size 1600

9. (15 points) Ages

A data scientist takes a random sample of 400 people in a large city. The ages of the sampled people have an average of 35 years and an SD (standard deviation) of 20 years.

The data scientist bootstraps the sample 10,000 times, calculates the mean age of each bootstrapped sample, and finds the interval that contains the middle 95% of the 10,000 bootstrapped means. The interval goes from 33 years to 37 years.

(a) (3 pt) The interval (33 years, 37 years) is an approximate 95% confidence interval for the _____ of the people in the _____.

Fill in the blanks above by selecting from the following options.

(i) Blank 1 (make **exactly one** choice):

○ ages	○ average age	○ average	
○ sample	○ sample mean	○ city	○ city mean

(ii) Blank 2 (make **exactly one** choice):

○ ages	○ average age	○ average	
○ sample	○ sample mean	○ city	○ city mean

(b) (3 pt) The distribution of the ages of the sampled people (pick **exactly one** option):

○ is approximately normal by the Central Limit Theorem.
○ is approximately normal, but not because of the Central Limit Theorem.
○ is not normal, not even approximately.
○ may be approximately normal, or not; we need more information to decide.

(c) (3 pt) True or false: Approximately 95% of the people in the sample are between 33 and 37 years old.

○ True ○ False

(d) (3 pt) True or false: Approximately 95% of the people in the city are between 33 and 37 years old.

○ True ○ False

(e) (3 pt) The city is in a country where the average age is 35.5 years. If possible, perform a statistical test of whether or not the average age in the city is 35.5 years, using 1% as the cutoff for the p-value. State your conclusion by picking **exactly one** of the options below.

○ Since the p-value cutoff and the confidence level of the interval are inconsistent, we cannot perform this test.

○ The test concludes that the data are consistent with the hypothesis that the average age in the city is 35.5 years.

○ The test concludes that the data are not consistent with the hypothesis that the average age in the city is 35.5 years.