

Problem 9

WINTER 24 FINAL

In 2024, the Olympics will include breaking (also known as breakdancing) for the first time. The breaking competition will include **16 athletes**, who will compete in a single-elimination tournament.

In the first round, all 16 athletes will compete against an opponent in a face-to-face "battle". The 8 winners, as determined by the judges, will move on to the next round. Elimination continues until the final round contains just 2 competitors, and the winner of this final battle wins the tournament.

The table below shows how many competitors participate in each round:

Round	Competitors
1	16
2	8
3	4
4	2

After the 2024 Olympics, suppose we make a DataFrame called `breaking` containing information about the performance of each athlete during each round. `breaking` will have one row for each athlete's performance in each round that they participated. Therefore, there will be $16 + 8 + 4 + 2 = 30$ rows in `breaking`.

In the `"name"` column of `breaking`, we will record the athlete's name (which we'll assume to be unique), and in the other columns we'll record the judges' scores in the categories on which the athletes will be judged (creativity, personality, technique, variety, performativity, and musicality).

"name"
p1
p2
p3
.
.
p16

p2
p3
p6

8 people whose name appears 1X

4 people whose name appears 2X

2 people whose name appears 3X

2 people whose name appears 4X

Problem 9.1

How many rows of `breaking` correspond to the winner of the tournament? Give your answer as an integer.

Click to view the solution.

1st place; appears 4 times
(one per round)
2nd place; also appears 4 times

Problem 9.2

How many athletes' names appear exactly twice in the `"name"` column of `breaking`? Give your answer as an integer.

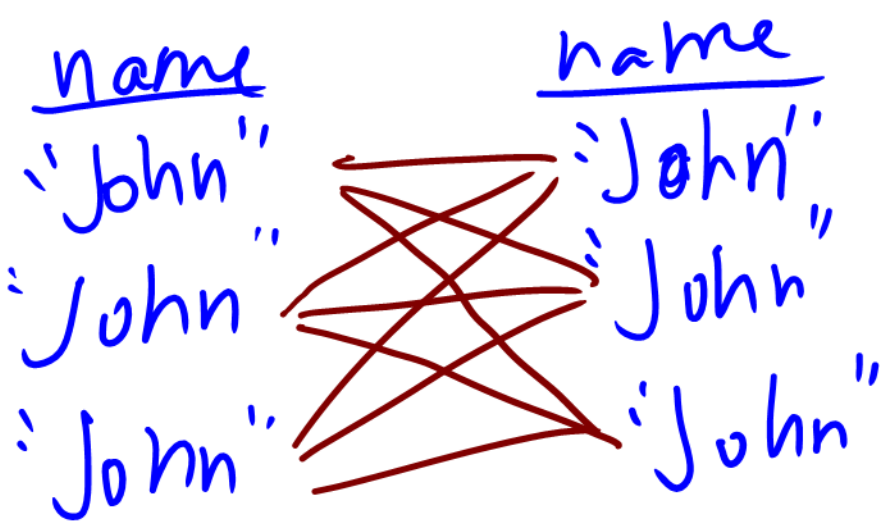
Click to view the solution.

4 (see above)

Problem 9.3

If we merge `breaking` with itself on the `"name"` column, how many rows will the resulting DataFrame have? Give your answer as an integer.

Hint: Parts (a) and (b) of this question are relevant to part (c).



1 row
 8 people whose name appears 1x
 4 people whose name appears 2x → 4 rows
 2 people whose name appears 3x → 2 rows
 2 people whose name appears 4x → 2 rows
 → 16 rows

$$8 \cdot 1 + 4 \cdot 4 + 2 \cdot 9 + 2 \cdot 16$$

$$8 \quad \quad \quad 16 \quad \quad \quad 18 \quad \quad \quad 32$$

$$\quad \quad \quad \underbrace{\quad \quad}_{24} \quad \quad \quad \underbrace{\quad \quad}_{50} \quad \quad \quad \textcircled{74}$$

Problem 13

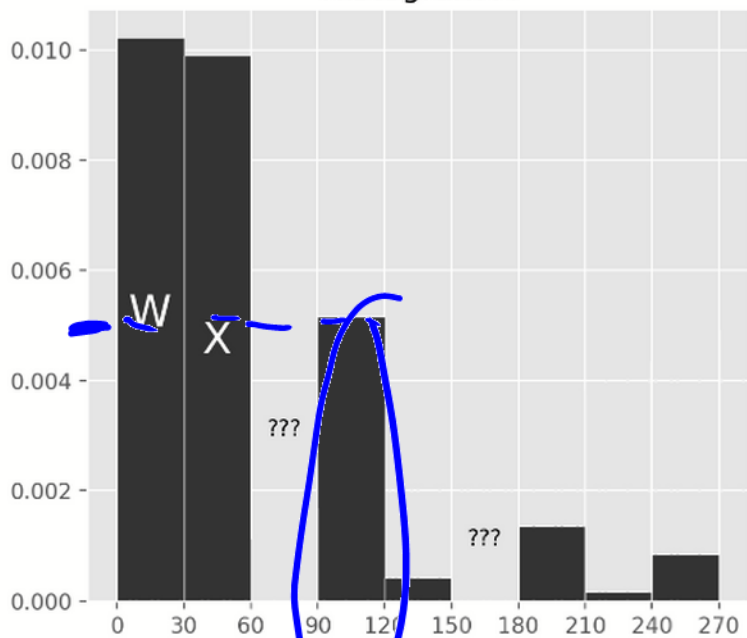
FALL 23 FINAL

Ashley doesn't have access to the entire `txn` DataFrame; instead, she has access to a simple random sample of 400 rows of `txn`.

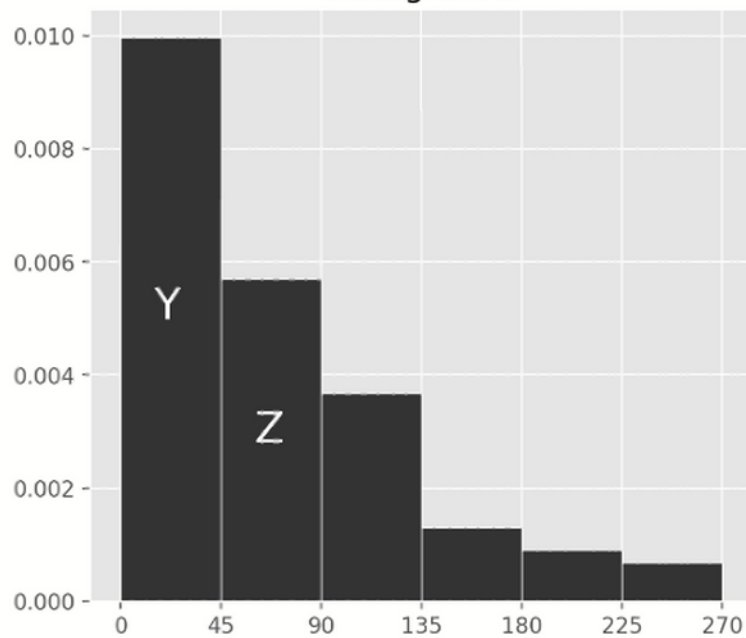
She draws two histograms, each of which depicts the distribution of the `"amount"` column in her sample, using different bins.

400 transactions each

Histogram A



Histogram B



which bin contains 60

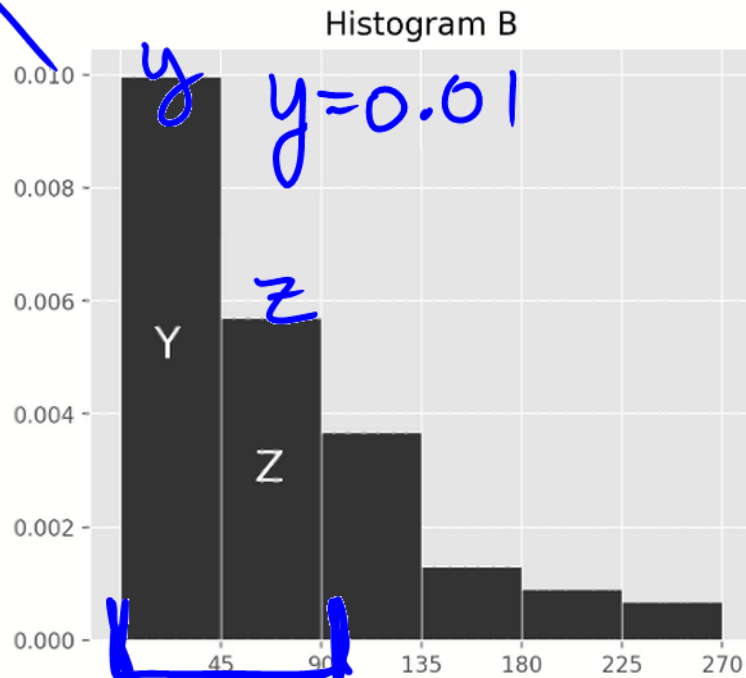
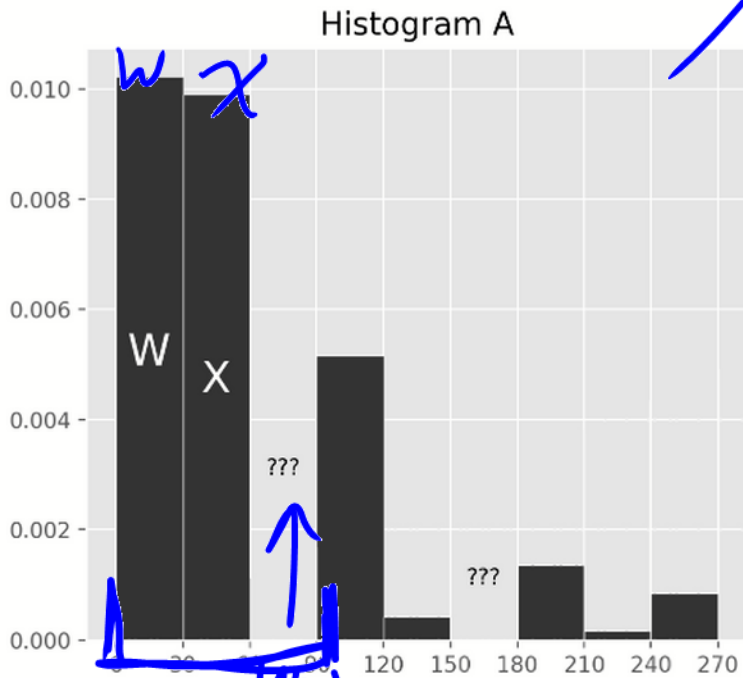
$$\frac{60}{400} = \frac{3}{20} = 0.15 \leftarrow \text{bin with this area}$$

$$30 \times h = 0.15$$

area

$$h = \frac{0.15}{30} = \frac{0.05}{10} = 0.005$$

same data



area of
first 3 bins
on left

= area of
first 2
bins on right

$$30 \cdot w + 30 \cdot x + 30 \cdot \underline{u} = 45 \cdot y + 45 \cdot z$$

$$30(w + x + u) = 45(y + z)$$

$$w + x + u = \frac{3}{2}(y + z)$$

$$u = \frac{3}{2}(y + z) - w - x$$