Problem 1   Students

The table below shows the distribution of rents paid by students in a college town. The first column consists of ranges of monthly rent, in dollars; bins include the left endpoint but not the right. The second column shows the percent of students who pay rent in each of the ranges.

Dollars	Student %
250-350	25
350-550	25
550-950	25
850-1350	25

- True or false (explain): The data show that the rents are evenly distributed over the interval 250-1350.
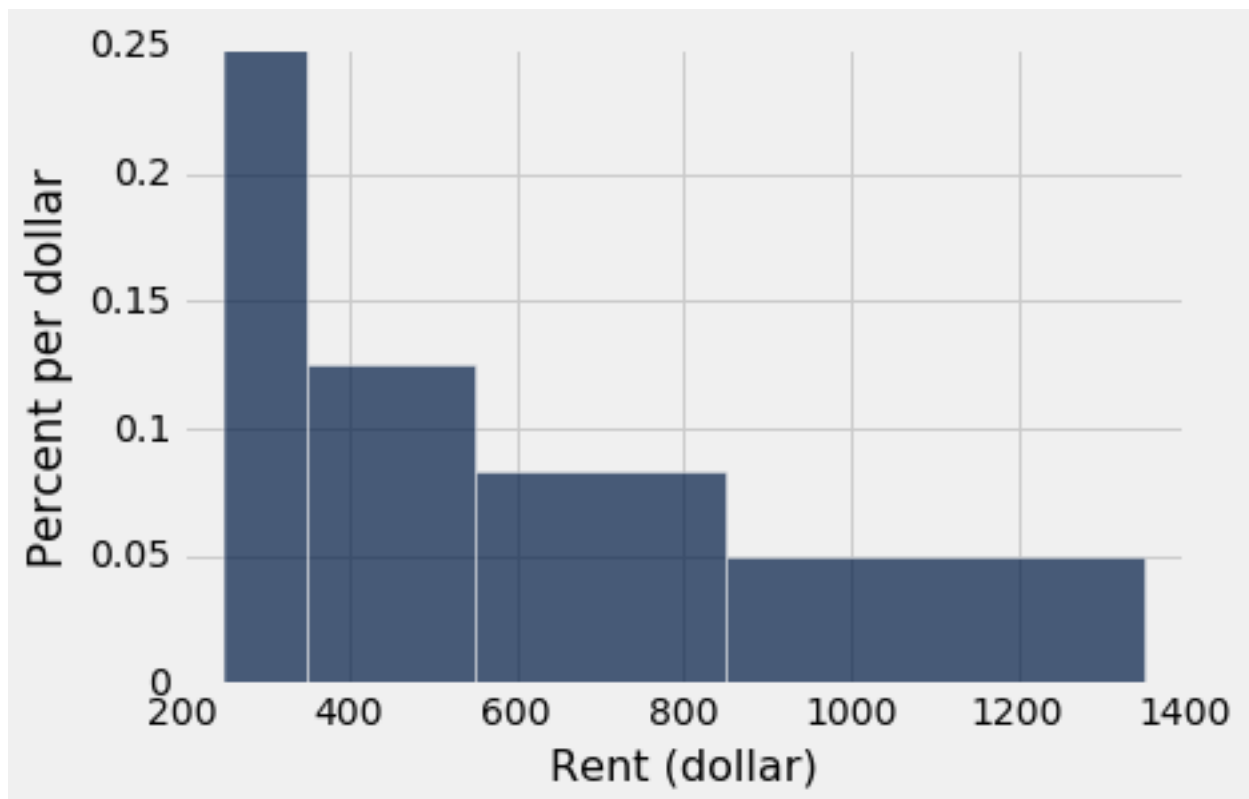
- Draw a histogram of the data

- What is the height and the correct units of the histogram bar over the bin 250-550 on the density scale?

  (i) 12.5% per student

  (ii) 0.125% per student

  (iii) 0.125% per dollar

  (iv) 12.5% per dollar

Answer:

(a) False. Each bin contains 25% of the rents, but bins don't all have the same width.

(b) **Note**: This is intended as a problem that can be done by hand on a midterm.


(c) (iii)

Problem 2   Edges

Students in a Data Science class are drawing a histogram of the distribution of daily maximum temperatures in their city. According to their data table, 7% of the temperatures are in the range 70 to 75 degrees Fahrenheit; intervals include the left endpoint but not the right. Two students are arguing about the height of the bar over the bin 70-75 degrees. One student says, "The bin is 5 degrees wide, so the height of the bar is $\frac{7\%}{5}$ degrees = 1.4% per degree." The other student says, "But the interval doesn't include the right endpoint, so it includes 74 but not 75. The bin is 4 degrees wide and the height is $\frac{7\%}{4}$ degrees = 1.75% per degree."

Who is right? Explain.

Answer: The first student. The bin doesn't end at 74 degrees temperatures such as 74.2 degrees, 74.6 degrees, 74.9999999 degrees are all in the bin. The width of the bin is 5 degrees.

Problem 3   Incomes
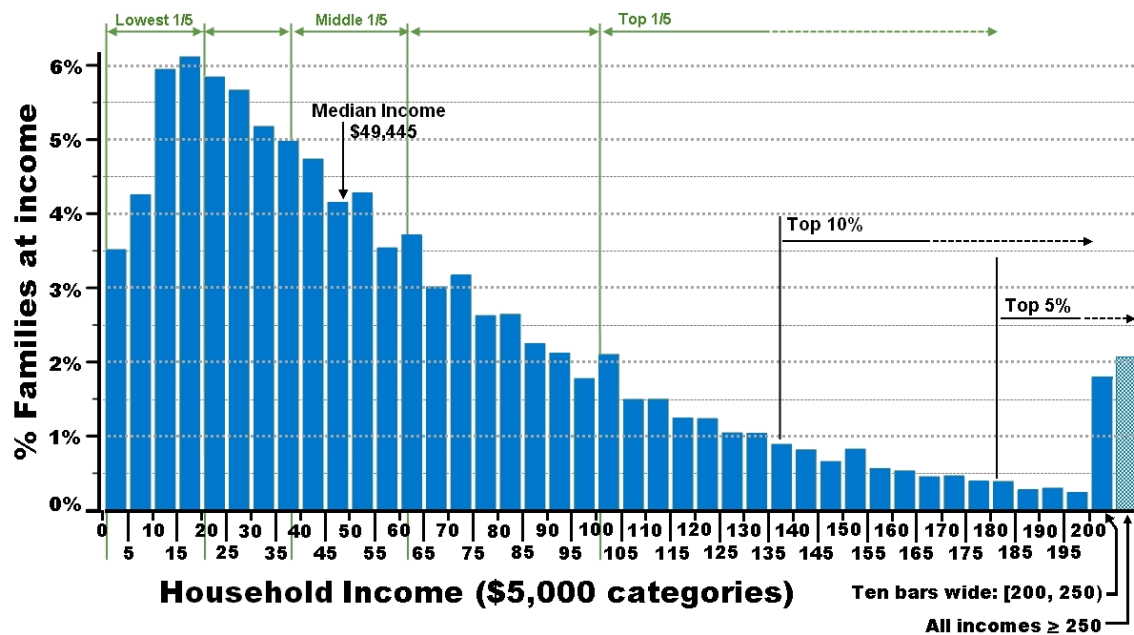
Data source: http://www.census.gov/hhes/www/cpstables/032011/hhinc/new06_000.htm
The figure above is based on Census data. Its purpose is to show the distribution of household income in the United States in 2010, among households that have annual incomes under $250,000.

(a) The top left of the figure indicates that the bottom 20% of the incomes (the "Lowest 1/5") are in the range $0 to $20,000. About what percent of all the households have incomes in the range $0 to $5,000?

(b) The vertical axis is labeled "% of Families at income." True or False (explain): About 6% of the households have an income rounded to the nearest $1,000 of $12,000.

(c) The figure is almost a histogram, but not quite. What would you have to do to turn it into a histogram?

Answer:

(a) The only way to understand the "Lowest 1/5" is to assume that the label on the vertical axis denotes the percent in each bin. So the percent in the first bin is 3.5%.

(b) False. 6% of the households have incomes in the range $10,000 to $15,000. If you assume that those households are about evenly distributed over the range, then the percent of households with incomes in the "$11,500 to $12,500" range is about $\frac{6\%}{5} = 1.2\%$. The answer would be the same for any range that is $1,000 wide and is contained in the range $10,000 to $15,000. If the assumption of even distribution is not justified, then these estimates will be off.

(c) The main points are to fix the labeling and scale on the vertical axis and take care of the two bars at the right end.