

Name: KEY

Date: _____ Period: _____

Without discussing with your classmates, answer the following questions—write one answer on each of your 12 slips of paper. Place each slip in the appropriate paper bag.

- What is your gender?
- How many text messages did you send yesterday?
- How many people under the age of 18 live in your house?
- Which is your favorite food from the following choices: pizza, hamburgers, sushi, salad, chicken, other?
- Do you have a job that pays by the hour?
- If yes, how many hours do you work in an average week? (If no, put 0.)
- List how many hours you work a week (again) and then the average number of hours that you study a week.
- List your gender (again) and your shoe size.
- List your gender (again) and number of text messages (again).
- What is your favorite kind of music?

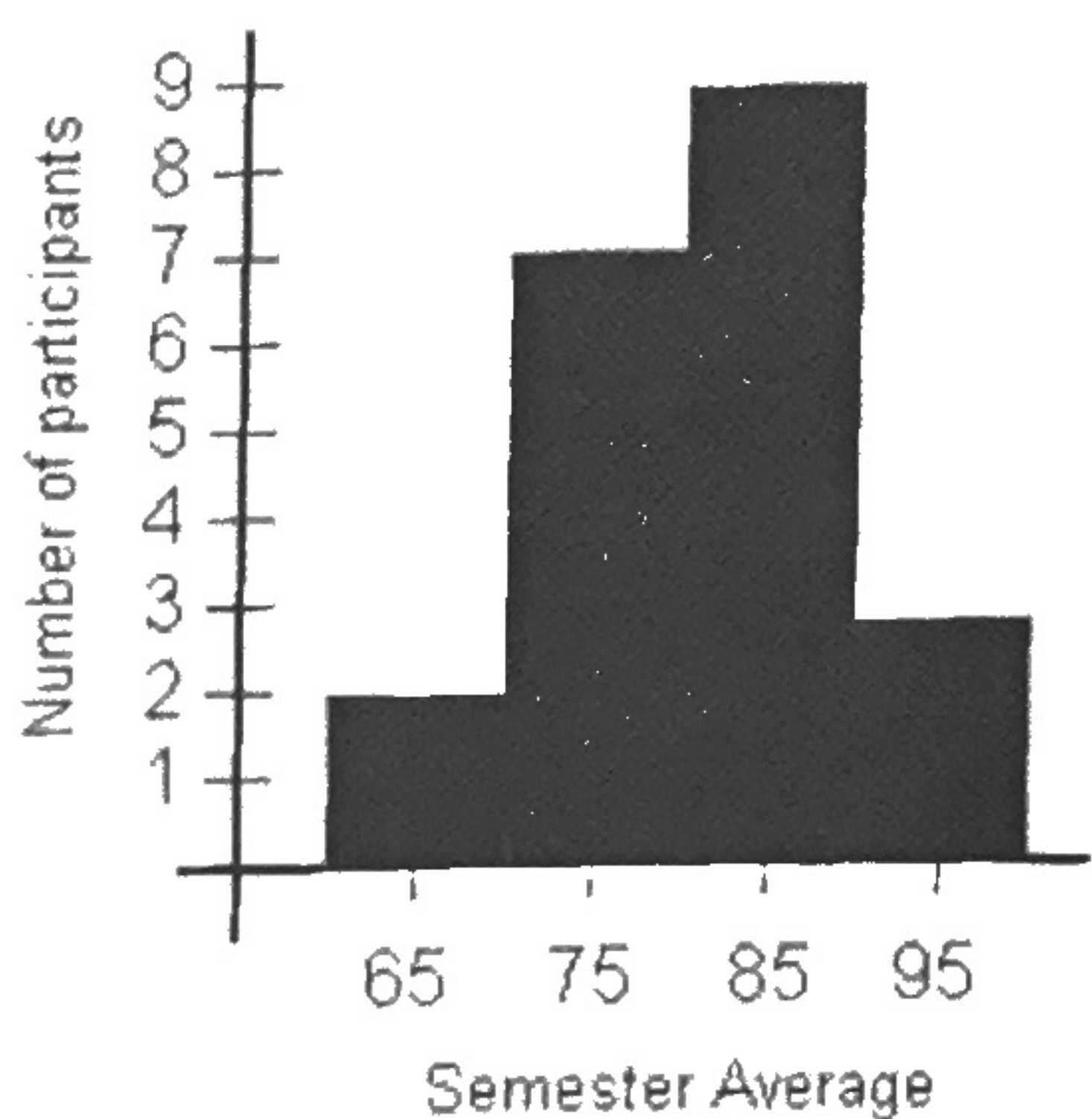
1. What is Categorical Data (Qualitative Data)?
 information about how study subjects fit into different categories.
 ex. male or female; blonde or brunette, etc.

2. What is Quantitative Data?
 information that is recorded numerically.
 ex. height, age, income.

3. What do you think **univariate** means?
 "one variable;" into about one variable is recorded

4. What do you think **bivariate** means?
 "two variables;" into ~~recorded~~ recorded about 2 variables.

Use the following histogram to answer Questions 5-7.



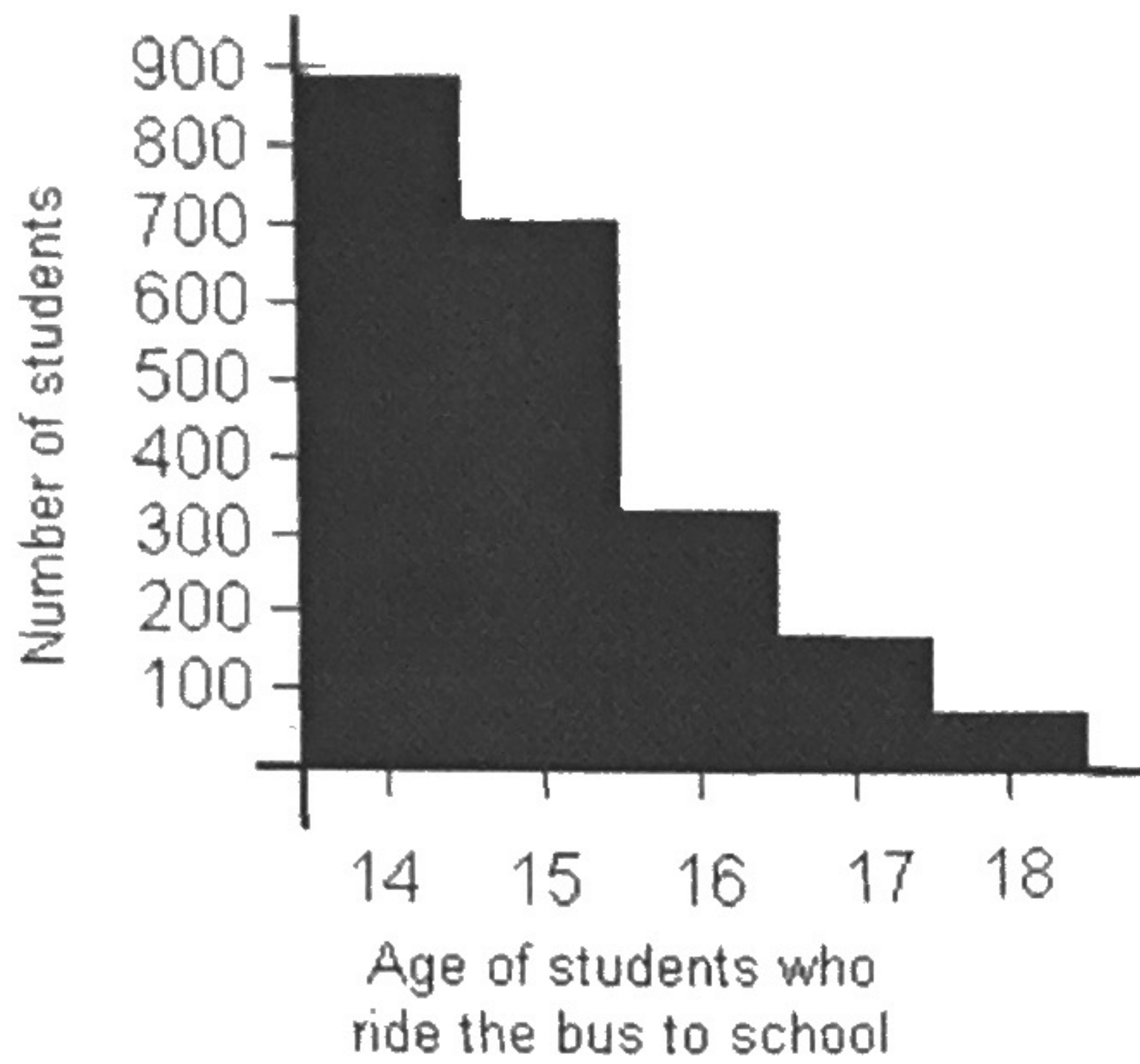
5. Does the histogram illustrate univariate data or bivariate data? Categorical or quantitative?
 univariate - semester average. since this is recorded numerically, this is quantitative.

6. Would you describe the histogram as symmetric? Why or why not?
 fairly symmetric since the two sides roughly reflect each other.

7. Describe the distribution of grades for the class. Justify any estimates you make.

Average is somewhere around 84(ish).
 you can approximate $\frac{2(65) + 7(75) + 9(85) + 3(95)}{21} = 81.2$
 may not all have the same grade²¹ though.

Use the following histogram to answer Questions 8 and 9.



8. Consider the new histogram showing bus ridership at a local high school. This histogram is not symmetric; it is **skewed to the right**. You know this because the distribution has **a tail out toward the right side**. What is happening with this population that causes the distribution to be skewed to the right?

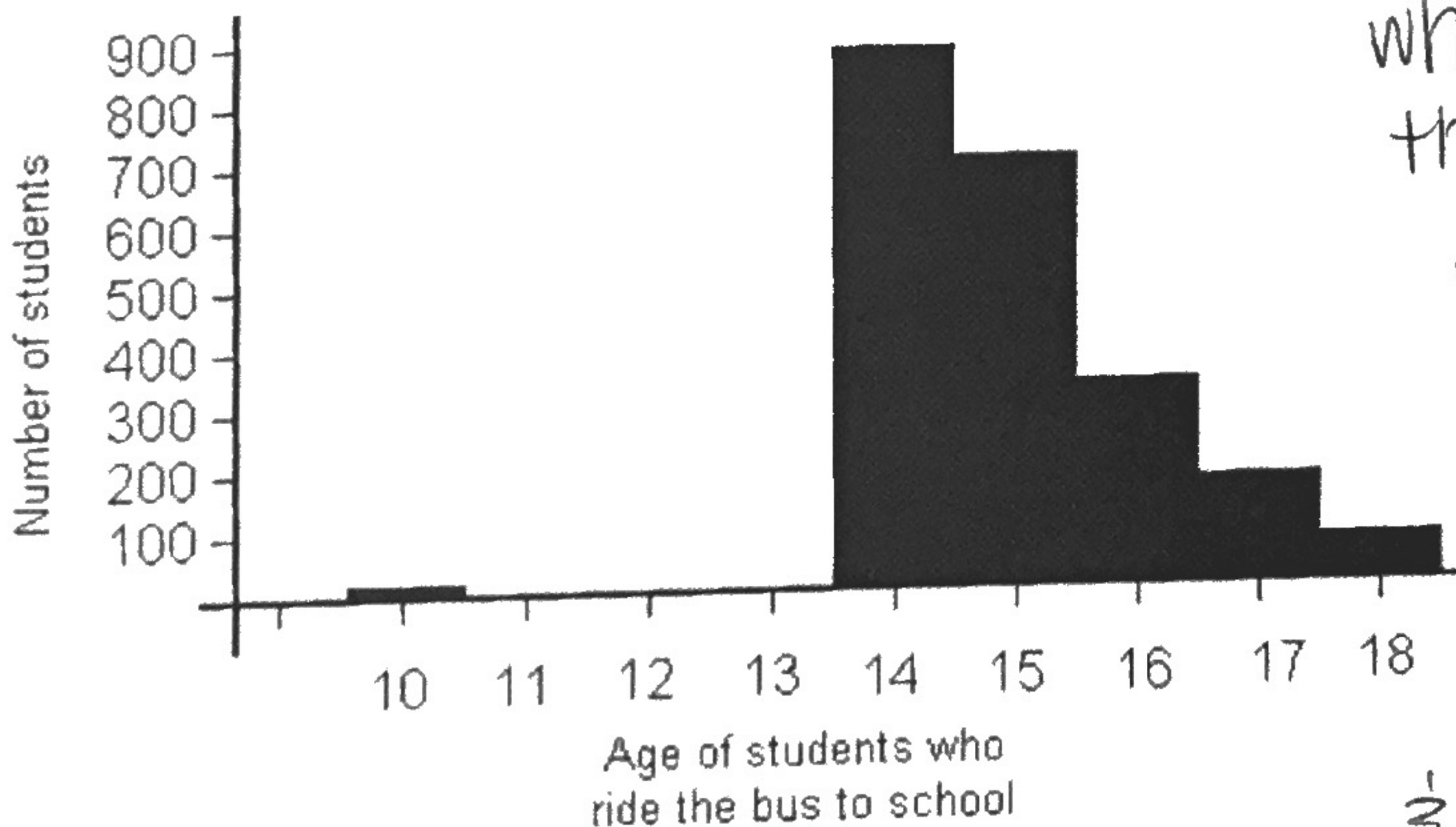
As the age of students who ride the bus increases, the # of students decreases. More students could be driving at 16, some 18yr olds may have graduated.

9. Estimate the average age of students who ride the bus. This average is the center of the distribution. Justify your estimate.

average age around 15(ish).

$$\text{or } \frac{14(900) + 15(700) + 16(300) + 17(200) + 18(100)}{2200} = 15.05 \text{ yrs}$$

10. Suppose the histogram of bus ridership looked like this instead. Data values that are distant from most of the other values are generally thought of as **outliers**. What may have happened? Does this data distribution affect your answers to Questions 8 and 9?

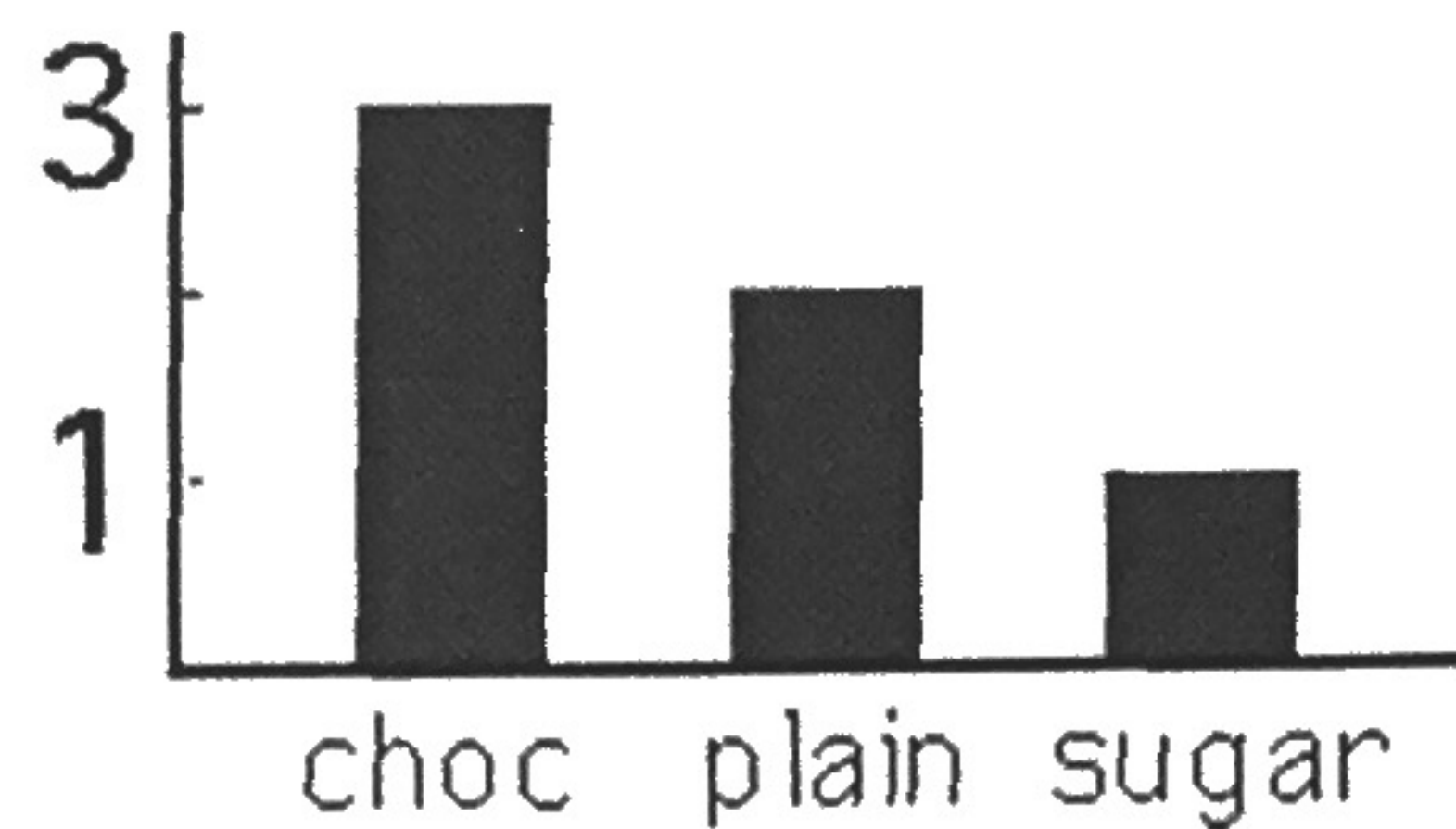


Either there's a 10-year old who skipped a grade or there was an error in data recording. you need to check to see if you need to correct your data. if it's true, you need to refigure the average age. $\frac{1}{2}$ note the outlier.

Now, how does a histogram compare to a bar graph?

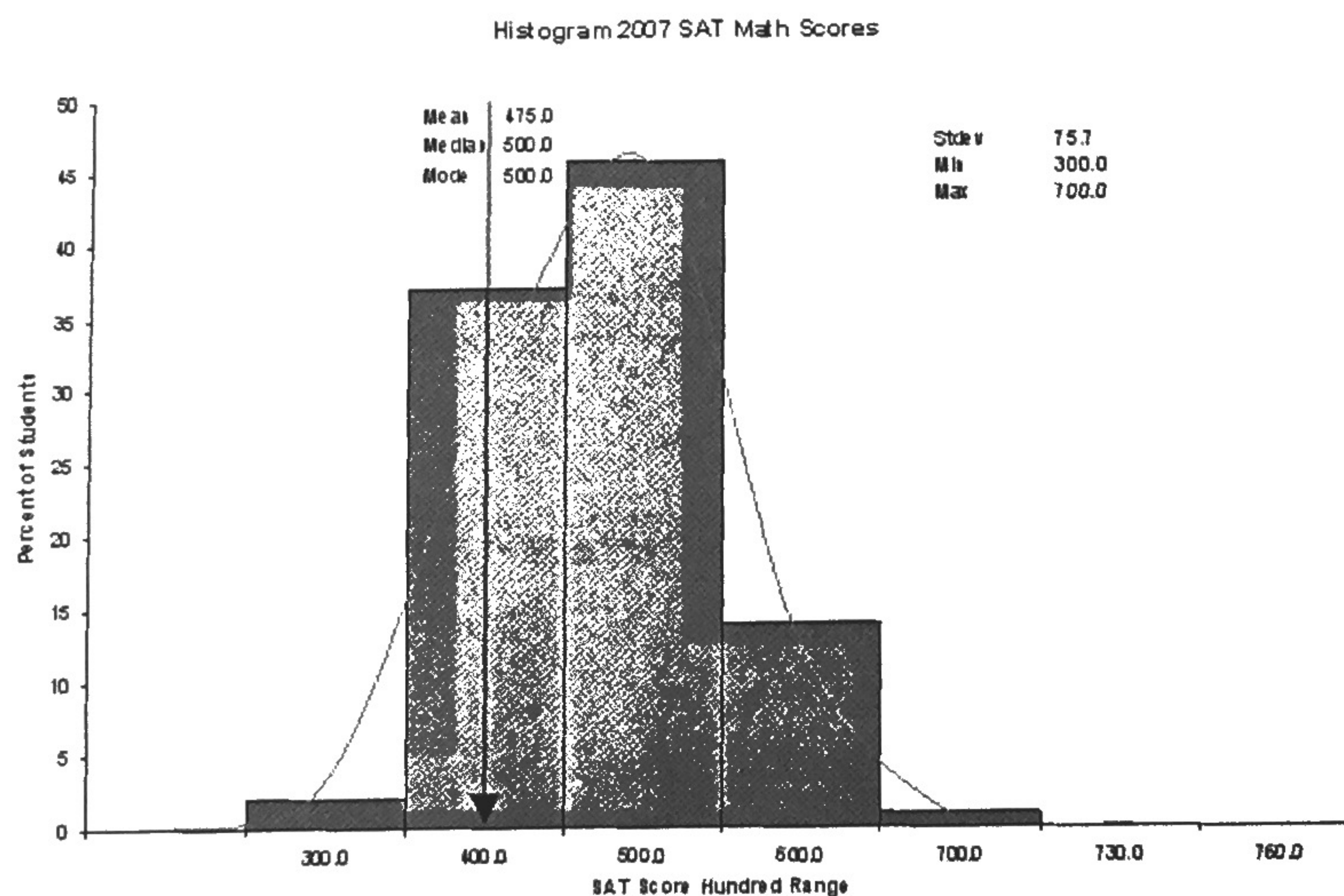
There are two differences, one is in the **type of data** that is presented and the other in the **way they are drawn**.

In **bar graphs** are usually used to display "**categorical data**", that is data that fits into categories. For example suppose that I offered to buy donuts for six people and three said they wanted chocolate covered, 2 said plain and one said with icing sugar. I would present this in a bar graph shown below:



Histograms, on the other hand, are usually used to present "**continuous data**", that is data that represents **measured quantity** where, at least in theory, the **numbers can take on any value in a certain range**.

A good example is SAT scores. The SAT is a test that many students take before they go to college and many colleges decide who to accept based on the SAT score. One part of the test focuses only on math and the test score ranges from 200 to a maximum of 800. Almost all students score over 300, most of the students score between 400 and 600, and the very best students score 700 or more. Here is a histogram of the percent of students taking the math SAT getting scores in each range of 100, from 300 to 700.



- The difference in the way that bar graphs and histograms are drawn is that the bars in bar graphs are usually separated where in histograms the bars are adjacent to each other.

Another good example of where a histogram would be helpful is when you're talking about weight. If you measure the weights of a group of adults you might get numbers between say 90 pounds and 240 pounds. We usually report our weights as pounds or to the nearest half pound but we might do so to the nearest tenth of a pound or however accurate the scale is.