

AP Statistics - Graphs of Quantitative Data and Shape - Lesson

Common Core AP STAT 1.4, AP STAT 1.5, HSS-ID.A.1, HSS-ID.A.3 · maths-pdf.fr/en/

A quantitative variable is shown with a dotplot, stemplot, or histogram, and described by shape, outliers or unusual features, center, and spread in context. Shape may be symmetric or skewed, unimodal or bimodal. A skewed distribution is named for its long tail, so a tail toward large values means skewed right.

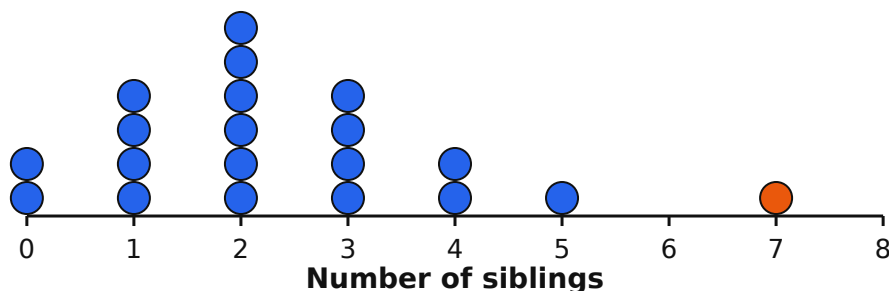
AP Statistics · Common Core State Standards: **AP STAT 1.4, AP STAT 1.5, HSS-ID.A.1, HSS-ID.A.3**

Key facts

- Dotplots and stemplots show every value; histograms group values into equal-width bins.
- Histogram bars touch because the horizontal axis is a numerical scale.
- Describe a distribution with shape, outliers or unusual features, center, and spread.
- A long tail to the right means skewed right; to the left means skewed left.
- Mention peaks, gaps, and clusters, and always answer in the context of the data.

How do you display a quantitative variable?

A quantitative variable needs a graph that shows the numerical scale. A **dotplot** places one dot above each value on a number line and works best for small data sets. A **stemplot** splits each value into a stem and a leaf. A **histogram** groups values into equal-width intervals and shows how many fall in each.



A dotplot puts one dot per student above its value. The value 7 is separated from the rest of the data.

```

6 | 2 8
7 | 1 3 5 5 8
8 | 0 2 4 4 5 7 8
9 | 0 1 4 7

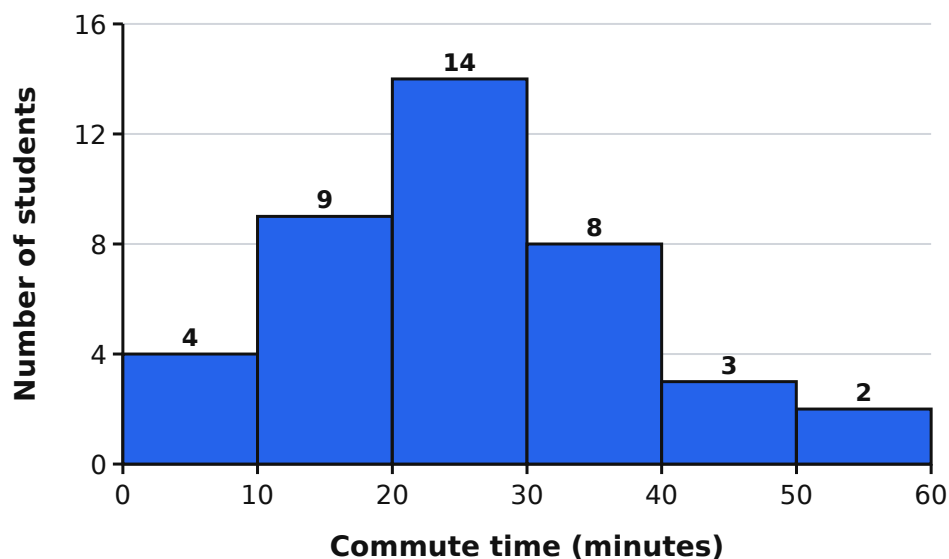
```

Key: 7|3 = 73 points

A stemplot keeps every value. The tens digit is the stem and the ones digit is the leaf.

Histogram rules

Use equal-width intervals, label both axes with units, and make the bars touch, because the horizontal axis is a number line. Choose a bin width that shows the pattern without too much detail.



Histogram of 40 commute times. Each bar covers an interval; a value on an edge goes in the upper interval.

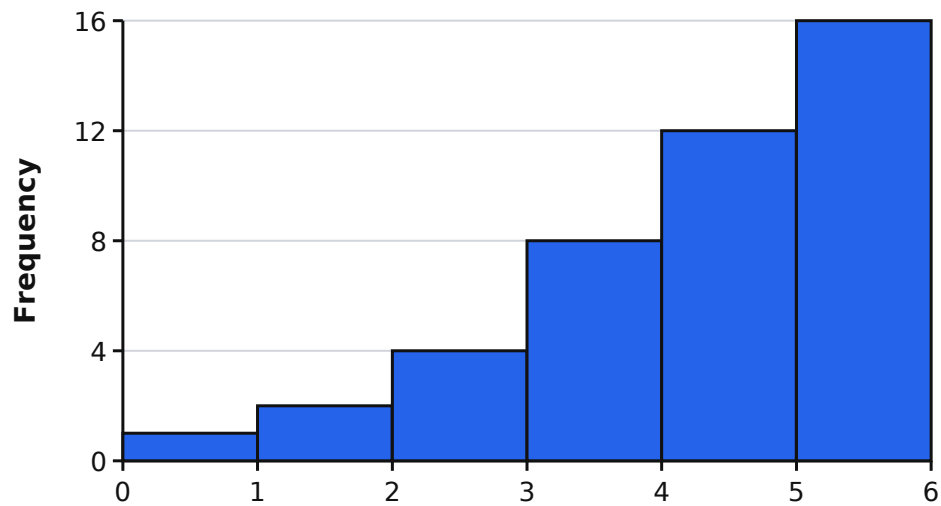
What do you say when you describe a distribution?

AP graders expect four features, in context, often remembered as SOCS: **shape**, **outliers** or other unusual features, **center**, and **spread**.

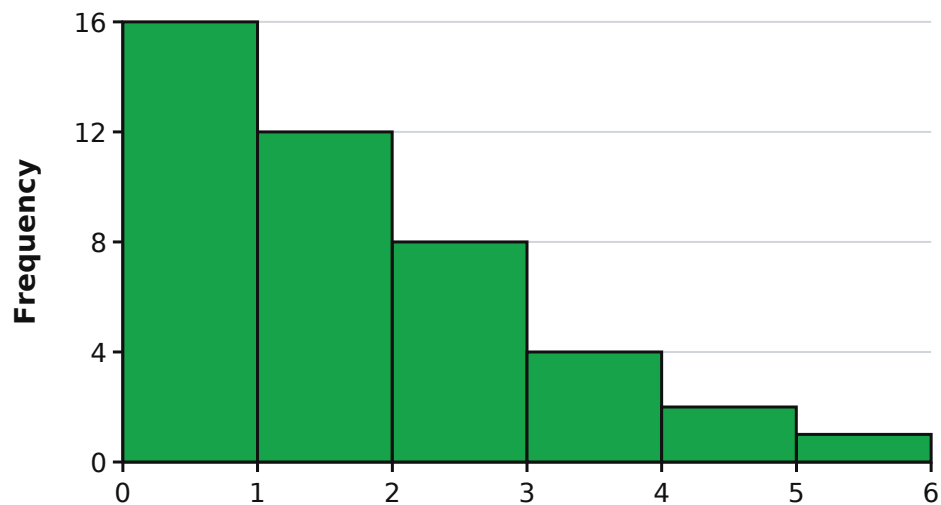
- **Shape:** symmetric, skewed right, skewed left, unimodal, bimodal, or uniform.
- **Outliers:** values far from the rest, plus any gaps or clusters.
- **Center:** a typical value, such as the median.
- **Spread:** how far the values range, such as the minimum to maximum.

How do you tell skewed from symmetric?

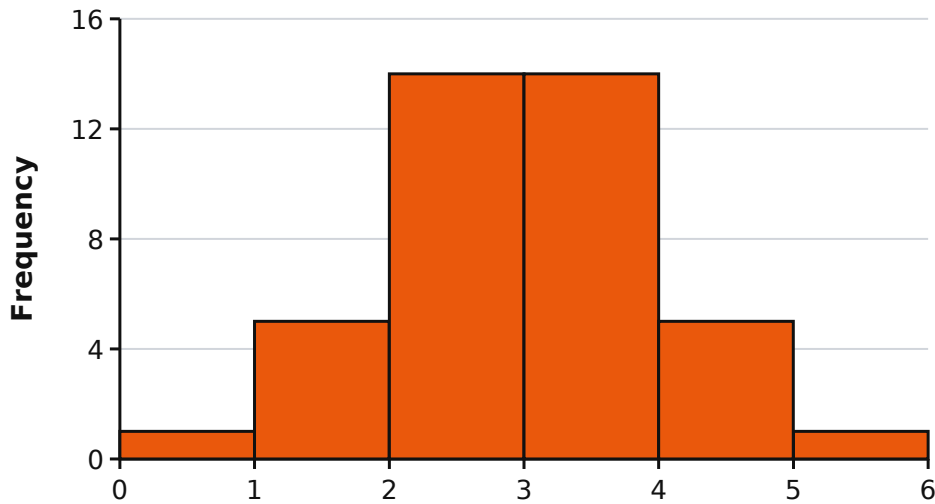
Skewness is named for the long tail. A distribution **skewed right** has a tail toward large values; **skewed left** has a tail toward small values.



Skewed left: the tail points to the lower values.



Skewed right: the tail points to the larger values.



Symmetric: the left and right sides are mirror images.

Worked example. In the commute histogram, 14 students travel 20 to under 30 minutes. The distribution is unimodal and skewed right, with a typical commute of about 25 minutes. Times range from 0 to 60 minutes with no gaps. Spread: $60 - 0 = 60$ minutes at most.

Questions about graphs of quantitative data and shape

What is the difference between a bar graph and a histogram?

A bar graph displays a categorical variable and its bars are separated. A histogram displays a quantitative variable on a number line, so its bars touch and each covers an interval of values.

Which interval does a value on a bin edge belong to?

By the usual convention, a value equal to an edge goes in the interval that starts at that edge. For example, 30 minutes belongs to the 30 to under 40 bar.

What does skewed right mean?

It means the right tail is longer, so there are a few unusually large values. The bulk of the data sits toward the lower values, as with incomes or commute times.

What is a bimodal distribution?

A bimodal distribution has two distinct peaks. This often signals two different groups mixed together, such as heights of men and women combined, and you should say so in context.

Why does the bin width matter?

Very wide bins hide features, while very narrow bins make the graph look jagged. A good

choice, typically 5 to 10 bins, shows the overall shape and any outliers.