Making a histogram

Twenty numbers in a row are hard to take in. How do you turn a list of numbers into bins and bars, and how wide should the bins be?

WORDS TO KNOW

Bin: a stretch of the number line, such as 20 to 30.

Edge: a number where one bin ends and the next begins.

Empty bin: a bin with a count of 0. It still keeps its place.

Bin width: how long each bin is. In one histogram every bin has the same width.

Frequency table: a list of the bins in order, with the count of values in each.

KEY IDEAS

1. From a list to a table

Pick the bin width. The first edge is a round number at or below the smallest value, usually a multiple of the width. Add bins until one holds the largest value. Then count the values in each bin.

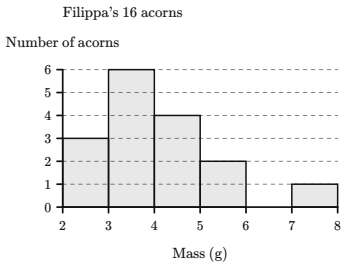
Edge rule: a value on an edge goes in the bar to its right. The bin "20 to 30" holds every x with $20 \le x < 30$.

So a largest value sitting exactly on an edge needs one more bin. With width 10 and a largest value of 40, the bins go on to "40 to 50".

Sorting the list first makes counting safe. Before you draw, check that the counts add to the number of values, and that every bin from first to last is listed, even one with a count of 0.

2. From the table to bars

The bottom axis is a number line: edges at equal spacing, named with the quantity and unit, such as "Mass (g)". It may start at the first edge, not at 0. The side axis counts values from 0.



Each bar covers exactly its bin and is as tall as its count. Neighboring bars touch. An empty bin gets no bar but keeps its place as a gap.

In Filippa's acorns, no acorn weighs from 6 g up to 7 g. The gap shows that the heaviest acorn is far from the rest.

3. The width is a choice

Bins that are too narrow give ragged bars with many accidental gaps. Bins that are too wide give two or three bars and hide where values crowd. About 5 to 15 bins is a common guide, fewer for small data sets.

$$\text{width} \approx \frac{\text{largest} - \text{smallest}}{\text{number of bins you want}}$$

Round **up** to a convenient width such as 1, 2, 5, 10, 20 or 0.5. Then count the bins you really get.

If a question asks "how many are at least 25?", choose bins so that 25 is an edge. Then the answer is a set of whole bars.

Free sample from math.wiki for classroom use. Do not post. math.wiki

WATCH OUT

- **Do not put an edge value in the bar to its left.** The counts still add to the total, so this mistake is easy to miss.
- **Do not skip an empty bin.** Leaving it out makes the number line jump and hides a gap.
- **The first edge can never be above the smallest value.** That value would have no bin.
- **Bins must all have the same width.** A wider bin collects more values just by being wide.

WORKED EXAMPLES

Example 1. Ana timed 14 bus rides, in minutes. 18, 7, 22, 13, 25, 10, 16, 37, 20, 14, 9, 19, 24, 15. Make a frequency table with bins 5 minutes wide.

1. The smallest value is 7. The largest multiple of 5 at or below it is 5, so the first edge is 5.
2. The largest value, 37, lies in "35 to 40". The edges run 5, 10, ..., 40, so there are $\frac{40-5}{5} = 7$ bins.
3. Sort: 7, 9, 10, 13, 14, 15, 16, 18, 19, 20, 22, 24, 25, 37. The values 10, 15, 20 and 25 sit on edges, so each goes in the bin it starts.

Ride (min)	5 to 10	10 to 15	15 to 20	20 to 25	25 to 30	30 to 35	35 to 40
Count	2	3	4	3	1	0	1

Answer: the table above. The bin 30 to 35 is empty and is drawn as a gap. **Check:** 2 + 3 + 4 + 3 + 1 + 0 + 1 = 14.

Example 2. Fredrik weighed 40 acorns. The lightest was 1.8 g and the heaviest 9.3 g. Choose a width that gives about 6 bins.

1. The spread is 9.3 − 1.8 = 7.5 g, and 7.5 ÷ 6 = 1.25.
2. Round up to a convenient width: 2 g.
3. The largest multiple of 2 at or below 1.8 is 0. The heaviest, 9.3, lies in "8 to 10".

Answer: width 2 g, with bins 0 to 2, 2 to 4, 4 to 6, 6 to 8 and 8 to 10. That is 5 bins, near the 6 planned.

Free sample from math.wiki for classroom use. Do not post. math.wiki